

C	10	34.8	11.0	15548	24	ABL34155	Human immune syste
C	11	34.8	10.9	1230025	20	AAG91090	Nucleotide sequenc
C	12	34.2	10.9	6169	22	AAS46370	Tumour suppressor
C	13	34.2	10.9	9741	24	ABL33323	Human immune syste
C	14	34.2	10.9	18624	24	ABL33703	Human immune syste
C	15	34	10.8	3088	12	AAO13727	Plasmodium falcipa
C	16	34	10.8	161425	22	AAH02340	Human AKAP10 gene
C	17	34	10.8	162025	22	AAH02339	Human AKAP10 gene
C	18	33.8	10.7	3492	18	AAV74345	Staphylococcus aur
C	19	33.8	10.7	5647	24	ABL33567	Human immune syste
C	20	33.8	10.7	5647	24	AAS61321	Human immune syste
C	21	33.8	10.7	6084	24	ABL33658	Human gene regulat
C	22	33.8	10.7	85680	21	AAF32299	BAC containing rep
C	23	33.8	10.7	611590	21	AAF22303	Arabidopsis thalia
C	24	33.6	10.7	217	18	AAV76716	Staphylococcus aur
C	25	33.6	10.7	5909	24	ABL33159	Human immune syste
C	26	33.6	10.7	8178	24	ABL32087	Human immune syste
C	27	33.6	10.7	8344	22	AAK68947	Human immune/haema
C	28	33.6	10.7	8344	22	AAK68948	Human immune/haema
C	29	33.6	10.7	17700	22	AAK68945	Human immune/haema
C	30	33.6	10.7	17869	24	ABL32104	Human immune syste
C	31	33.4	10.6	5675	24	ABL33276	Human immune syste
C	32	33.4	10.6	10616	22	AAK85565	Human immune/haema
C	33	33.4	10.6	10616	22	AAK85567	Human immune/haema
C	34	33.4	10.6	140035	24	AAS98600	Human genomic DNA
C	35	33.2	10.5	5957	22	AAK31548	Human DNA for a no
C	36	33.2	10.5	32190	22	AAU36479	Human musculoskele
C	37	33.2	10.5	38342	22	AAS46745	Tumour suppressor
C	38	33	10.5	719	22	AAH07311	Human cDNA clone (
C	39	33	10.5	2319	23	ABL21210	Drosophila melanog
C	40	33	10.5	10749	20	AAX30257	Borrelia burgdorfe
C	41	33	10.5	23439	18	AAV74349	Staphylococcus aur
C	42	32.8	10.4	9965	24	ABL33527	Human immune syste
C	43	32.6	10.3	2351	23	ABL18630	Drosophila melanog
C	44	32.6	10.3	3001	21	AAH51761	Chromosome 13q31-q
C	45	32.6	10.3	6255	24	ABL34604	Human metastasis a

ALIGNMENTS

RESULT		1
ID	AA229725	standard; DNA; 315 BP.
XX	AA229725:	
XX	27-MAR-2000	(first entry)
DT		
DE	Human lung specific gene-2.	
KW	Lung Specific Gene; lscg; human; diagnostic marker;	
KM	prognosticate; Lung cancer; diagnosis; ds.	
OS	Homo sapiens.	
XX		
PN	MO9960160-A1.	
PD	25-NOV-1999.	
PX		
PF	12-MAY-1999;	99WO-US10344.
PR	21-MAY-1998;	98US-0086212.
PA	(DIAD-) DIADEXUS LLC.	
PI	Yang F, Macina RA, Sun Y;	
PT	WPI; 2000-116320/10.	
PS	A new method for diagnosing, monitoring and staging lung cancer - Example 1; Page 35; 40pp; English.	

sequence 315 BP; 108 A; 58 C; 49 G; 100 T; 0 other;

```

conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

RESULT	2
AAV21209	
ID	AAV21209 standard; DNA; 1664976 BP.
XX	

PT complete genome sequence of methano-genic archaeon, *Methanococcus jannaschii* - useful in identification of *M. jannaschii* genome

The present sequence represents the complete 1.66-megabase pair genome sequence of the *Methanococcus jannaschii* strain.

sequence 16649/b BP; 568133 A; 264649 C; 258701 G; 573392 T; 101 other;

Sequence	Conservative	Mismatches	Indels	Gaps
174	0	65	0	0

RESULT 3  
ABL12244/c  
ID ABL12244 standard; cDNA; 5516 BP  
XX

23-MAR-2001; 2001WO-US09231.

23-MAR-2000; 2000US-191637P;  
11-JUL-2000; 2000US-0614150.

(PEKE ) PE CORP NY.  
Venter JC, Adams M, Li PMD, Myers EW  
WPI; 2001-6556860/75.

D P-RSDB; ABB68141.  
 XR  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from *Drosophila* and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XR  
 PS  
 XX Claim 1; SEQ ID NO 31214; 21np + Sequence Listing; English.  
 CC  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB101840-AB116175), expressed DNA  
 CC sequences (AB101840-AB116175) and the encoded proteins  
 CC (AAB571737-AB872072).  
 CC  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/publised\_pcl\_sequences.  
 CC  
 XX Sequence 5516 BP; 1631 A; 1248 C; 1166 G; 1471 T; 0 other;  
 XQ

Query Match	11.4%	Score 35.8;	DB 23;	Length 5516;
Best Local Similarity	51.6%;	Pred. No. 5.8;		
Matches 82; Conservative	0;	Mismatches 77;	Indels 0;	Gaps 0;

RESULT	4
ABLI12246/c	
ID	ABLI12246 standard; cDNA; 5999 BP.
XX	
AC	ABLI12246;
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster expressed polynucleotide SEQ ID NO 31220.
XX	
KW	Drosophila; developmental biology; cell signalling; insecticide;
XX	pharmaceutical; gene; ss.
OS	Drosophila melanogaster.
XX	
PN	WO200171042-A2.
PT	
PD	27-SEP-2001.
XX	
PF	23-MAR-2001; 2001WO-US09231.
XX	
PR	23-MAR-2000; 2000US-191637P.
PR	11-JUL-2000; 2000US-0614150.
XX	
PA	(PEKE ) PE CORP NY.
XX	
PI	Venter JC, Adams M, Li PWD, Myers EW;
XX	
DR	WPI; 2001-656860/75.
DR	P-PADB; ABB68143.
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell

interactions -

Claim 1: SEQ ID NO 31220; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from *Drosophila*. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABR16176-ABR3511), expressed DNA sequences (ABR01840-ABR16175) and the encoded proteins (ABR57737-ABR72072).

The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published.pdf](http://wipo.int/pub/published.pdf) sequences.

Query Match	11.48;	Score 35.8;	DB 23;	Length 5999;
Best Local Similarly	51.68;	Pred. No. 5.9;		
Matches 82; Conservative	0;	Mismatches 77;	Indels 0;	Gaps 0;

RESULT 5  
 ABA03041  
 ID ABA03041 standard; DNA; 2944528 BP.  
 XX  
 AC ABA03041;  
 XX  
 DT 05-FEB-2002 (first entry)  
 XX  
 DE *Listeria monocytogenes* EGD-e genome sequence.  
 XX  
 KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 KW vitamin B12; bacterial infection; disease; ds.  
 XX  
 OS *Listeria monocytogenes*.  
 XX  
 PN MO200177335-A2.  
 XX  
 PD 18-OCT-2001.  
 XX  
 PE 11-APR-2001; 2001WO-FR01119.  
 XX  
 PR 11-APR-2000; 2000FR-0004629.  
 XX  
 RA (INSP ) INST PASTEUR.  
 XX  
 XX Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsalhi H, Dehoux P,  
 PI Dusserret O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P,  
 PI Daniels J, Goebel W, Krefft J, Kuhn M, Ng E, Vazquez-Boland JA;  
 PI Dominguez-Bernal G, Garrido-garcia P, Tierrez-Martinez A, Amend A,  
 PI Chakraborty T, Dommann E, Hain T, Berche P, Charblt A, Durant L,  
 PI Perez-diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N,  
 PI Madueno E, De Pablos B, Wehlman J, Kaerst U, Entian K, Hauf J;  
 PI Rose M, Voss H;  
 XX  
 DR WPI: 2002-010914/01.  
 XX  
 PP Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment

The present sequence is the genome sequence of *Listeria monocytogenes* BGD-e. This sequence and fragments of this sequence are useful for selecting probes and primers for detecting genes in *L. monocytogenes* related organisms, and to study genetic polymorphisms and other genomes. Proteins (AB547297-AB550149) expressed from the present sequence are useful for raising specific antibodies. Identification of *L. monocytogenes* and related organisms, and for biosynthesis and biodegradation, especially biosynthesis of Vitamin B12. This sequence and proteins encoded by it are also useful for selecting compounds that regulate gene expression and cell replication and modulate *L. monocytogenes*-related diseases. In addition, this sequence and proteins encoded by it are useful in pharmaceutical and vaccine compositions for the treatment or prevention of infections by *L. monocytogenes* and related organisms.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [http://wipo.int/pub/published\\_pct/sequences](http://wipo.int/pub/published_pct/sequences).

Query Match	11.28;	Score 35.2;	DB 24;	Length 2944528
Best Local Similarity	60.48;	Pred. No. 30;		
Matches	58;	Conservative	0.0	0.0

bacaaatgaataaacgagaaatcgaatgaatcagtcttc 62

b 338711 cgaaccactcgaataaaacatcacctatgagccgc

## RESULT 6

AAI58172;

Human polynucleotide SEQ ID NO 375.

Alzheimer's; Parkinson's disease; Huntington's disease; amytrophic lateral sclerosis; neurodegeneration; central nervous system; CNS; amyotrophic lateral sclerosis; neurodegeneration; central nervous system; CNS;

leukaemia; ss. drug screening; arthritis; inflammation;

W0200153312-A1.

26-DEC-2000; 2000WO-US34263.

09-JUL-2000; 2000US-0598042.  
19-JUL-2000

19-OCT-2000; 2000US-0693036.  
29-NOV-2000; 2000US-07211

49-NOV-2000; 2000US-0727344.

7

PI Tang YT, Liu C,  
PI Wang Y, et al.

WPI; 2001-442253/47.  
E-PSDB: AAM39015  
DR  
DR

such as central nervous system injuries -

The invention relates to human nucleic acids (AA157798-AA161369) and the encoded polypeptides.

in gene therapy. A composition containing at least one of the polynucleotides are useful

system, such as peripheral nervous injuries, peripheral nervous

Alzheimer's, Parkinson's disease, Huntington's disease, lateral sclerosis, and central nervous system diseases, such as

utilisation of the activities such as: Immune system suppression

and catabolic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, antibody

4. Infections and inflammation, leukaemias and C.N.S disorders.

XXXXXX present did not form part of the printed specification.

sequence 259/ BP; 733 A; 607 C; 690 G; 567 T; 0 other;

Matches	68;	Conservative	0;	Mismatches	55;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	-----	--------	----	------	----

-----  
 cccccccgctcagatcgaagatgacacatagccctgtcctaattgaa 1493  
 ccc

...caccagcagaacgaattcatcgttagccgcattccc 1553  
360 ++ + 370

Scanned; DNA; 20811 Bp.

[illegible]

epidemic relapsing fever; endemic relapsing fever; Lyme disease; infection

W09858943-A1

30-DEC-1998.

18-JUN-1998; 98WO-1151 276A



PR 03-SEP-1997; 97US-0057483.  
 PR 20-JUN-1997; 97US-0050359.  
 PR 22-JUL-1997; 97US-0053344.  
 PR 22-JUL-1997; 97US-0053377.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (MEDI-) MEDIMMUNE INC.  
 XX  
 PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;  
 PI White OR;  
 XX  
 DR WPI; 1999-081217/07.  
 XX  
 PT New isolated Borrelia burgdorferi nucleic acids - used to develop  
 PT products for the detection, diagnosis, characterisation, prevention  
 PT and therapy of infections, particularly Lyme disease  
 XX  
 PS Claim 1; Page 851-867; 1128pp; English.  
 XX  
 CC AAX20248 to AAX20402 represent polynucleotide sequences isolated from  
 CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for  
 CC the detection, diagnosis, characterisation, prevention and therapy of  
 CC Bb infections, e.g. Lyme disease. They can also be used for the  
 CC production of biosynthetic products, e.g. enzymes. Borrelia belongs  
 CC to a family of motile, spiral-shaped bacteria called Spirochetes.  
 CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and  
 CC endemic relapsing fever, and Lyme borreliosis, more commonly known as  
 CC Lyme disease.  
 XX  
 SQ Sequence 26811 BP; 8978 A; 3456 C; 3596 G; 10780 T; 1 other;

Query Match 11.1%; Score 35; DB 20; Length 26811;  
 Best Local Similarity 49.7%; Pred. No. 13;  
 Matches 89; Conservative 0; Mismatches 90; Indels 0; Gaps 0;  
 QY 122 tcagttgtgtgctgtgtgagtcgccacattatcagaagtagtactacaattt 181  
 DB 14519 taactttttatcataattttacataagcacccttacaagttttataattt 14578  
 QY 182 tgaataatagattgcaacaataaactggaatttalygaacatcagtagaagaaatc 241  
 DB 14579 tgccataattatataataatgaataattcacaattacgtttaaagcacttaa 14638  
 QY 242 aacattccctcccttaccagagatcatttacttgcacacagagataattgtcagtgt 300  
 DB 14639 attattctaaataaataatgacatttaccattaaagaataaattgactttatat 14697

RESULT 8  
 ABL17122  
 ID ABL17122 standard; DNA; 4084 BP.  
 XX  
 AC ABL17122;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 2839.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ds.  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX

PA (PEKE ) PE CORP NY.  
 XX  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 PI  
 XX WPI; 2001-656860/75.  
 DR  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1; SEQ ID NO 2839; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
 CC sequences (AB101840-AB16175) and the encoded proteins  
 CC (AB57737-AB572072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
 XX  
 SQ Sequence 4084 BP; 1059 A; 913 C; 872 G; 1240 T; 0 other;

Query Match 11.0%; Score 34.8; DB 23; Length 4084;  
 Best Local Similarity 48.1%; Pred. No. 10;  
 Matches 99; Conservative 0; Mismatches 107; Indels 0; Gaps 0;  
 QY 83 cttagagatctacagctccctttagggagacatagaatcggtgtgtgtgtgtgtg 142  
 DB 3493 cactacatacatacatacatatataatgatacacaatgtaaaactatctcgttgaag 3552  
 QY 143 agtccacccttatctcaagttagtgaactcaaatcttgaaatagattgcaacaa 202  
 DB 3553 ctgccacacatttgcagcccaatgctcatcacaattgttacaacattatccaataa 3612  
 QY 203 taaactggaattatggaacacatcagtagaaggaataacaacattccatcccttaca 262  
 DB 3613 catgttttattatcttctcatcttgcacacatcaataagaacattgtttgatttca 3672  
 QY 263 gatcatttacttgcacactcagatata 288  
 DB 3673 gtcttttttctgtatctgtcatga 3698

RESULT 9  
 ABL32774/C  
 ID ABL32774 standard; DNA; 15122 BP.  
 XX  
 AC ABL32774;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Human immune system associated gene SEQ ID NO: 747.  
 XX  
 KW Human; immune system disease; cytosine methylation; antislutmatic;  
 KW antiarteriosclerotic; anti-naemic; cytosatic; nootropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200200928-A2.  
 XX  
 PD 03-JAN-2002.  
 XX

PD	03-JAN-2002.
XX	
PF	02-JUL-2001; 2001WO-EP07537.
XX	
PR	30-JUN-2000; 2000DE-1032529.
XX	
PR	01-SEP-2000; 2000DE-1043826.
XX	
PA	(EPIG-) EPIGENOMICS AG.
XX	
PI	Olek A, Piepenbrock C, Berlin K;
XX	
DR	WPI; 2002-130909/17.
XX	
PT	Nucleic acid comprising fragment of chemically modified gene, useful
PT	for diagnosis and treatment of diseases associated with abnormal
XX	cytosine methylation
XX	
PS	Claim 1; SEQ ID NO 2128; 32pp + Sequence Listing; German.
XX	
CC	The present invention provides a number of human immune system associated
CC	genes which are modified by the methylation of cytosines. The sequences
CC	CC genes which are modified by the methylation of cytosines. The sequences
CC	including eye diseases such as retinopathy, neovascular glaucoma and
CC	macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC	leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC	rheumatoid arthritis, psoriasis and infestations

Sequence 15548 BP; 4209 A; 247 C; 2903 G; 8189 T; 0 other;

Query March	11.0%;	Score 34.8;	DB 24;	Length 15548;
Best Local Similarity	55.9%;	Pred. No. 13;		
Matches 66;	Conservative 0;	Mismatches 52;	Indels 0;	Gaps 0
QY 152	tttatatcaagtagtgatgatactacaagaatttggaaaatagattgcacacaataactgga	211		
Db 14979	TTTATTATTAATATTAATAAAAAAATATTAATAAATCTTTAAAAAATAAAAAAATAAAAAAATAAAAAA	14920		
QY 212	gtttatcggaacatcagtcagtagaagaataacacatccatcccttcagagatcatt	269		
Db 14919	AATTATCATATATATTTAAAAAAATCTAATTAAACTTTTCATCATTTTAAACCTATTAT	14862		
RESULT 11				
AAx91990/C				
AAx91990 standard; DNA; 1230025 BP.				

	RESULT	11
	AA91990/c	
ID	AA91990 standard; DNA; 1230025 BP.	
XX		
AC	AA91990;	
XX		
DT	13-SBP-1999 (first entry)	
XX		
DE	Nucleotide sequence of the complete genome of Chlamydia pneumoniae.	
KW	Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis	
KW	sinsitis; purulent otitis media; erythema nodosum; pharyngitis;	
KW	vaccine; neutralising epitope; ss.	
XX		
OS	Chlamydia pneumoniae.	
PN	WO9927105-A2.	
PPN		
IPD	03-JUN-1999.	
XXX		
PPE	20-NOV-1998; 98WO-IB01890.	
XXX		
PER	04-NOV-1998; 98US-0107078.	
PER	21-NOV-1997; 97ER-0014673.	
XXX		
PA	(GEST ) GENSET.	
XXX		

DR WPI; 1999-357842/30.  
 XX  
 PT Genome sequence of Chlamydia pneumoniae  
 XX  
 PS Claim 1; Page 291-611; 1912pp; English.  
 XX  
 CC The present sequence represents the complete genome of Chlamydia  
 CC pneumoniae, and encodes proteins AYA34584-Y35879. C. pneumoniae causes  
 CC respiratory disease such as pneumonia and bronchitis and is thought  
 CC to be a contributing factor in heart disease, sarcoidosis, sinusitis,  
 CC purulent otitis media, erythema nodosum or pharyngitis. The polypeptides  
 CC encoded by the open reading frames of the C. pneumoniae genome (see  
 CC AYA34584-Y35879) can be used in immunogenic compositions as vaccines.  
 CC Vectors containing C. pneumoniae nucleotide sequences can also be  
 CC used as immunogenic compositions, especially where the vector directs  
 CC the expression of a neutralising epitope of C. pneumoniae.  
 XX  
 SQ Sequence 1330025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;  
 Query Match 11.0%; Score 34.8; DB 20; Length 1230025;  
 Best Local Similarity 52.0%; Pred. No. 33;  
 Matches 78; Conservative 0; Mismatches 72; Indels 0; Gaps 0;  
 QY 70 cttaaatatactcttagagatctacagcctcccttaggggacatacaagtcagtgt 129  
 Db 1072911 CTTAAGTTAGTAAATCTCTATCTAATCGACTTTCGTATCCCTTAAGAAAGAGCT 1072852  
 QY 130 gttgccttggtagtcacacattatcaagtagtactactcaaatlttgaaata 189  
 Db 1072851 GTTGTATTTTAAAGTCTTTTAAACAACACTTGAGTTGAGCCAAACAGTAGAAGA 1072792  
 QY 190 gattgcacacataaactgagttatgg 219  
 Db 1072791 AATTTCTCTCAAAAAAATTTGTTATAG 1072762  
 RESULT 12  
 AAS46370/C  
 ID AAS46370 standard; DNA; 6169 BP.  
 AC AAS46370;  
 XX  
 DT 18-DEC-2001 (first entry)  
 DE Tumour suppressor gene derived chemically modified sequence #92.  
 XX  
 KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic;  
 KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;  
 KW cytosine methylation; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200168912-A2.  
 XX  
 PD 20-SEP-2001.  
 XX  
 PF 15-MAR-2001; 2001WO-EP02955.  
 XX  
 PR 15-MAR-2000; 2000DE-1013847.  
 PR 06-APR-2000; 2000DE-1019058.  
 PR 07-APR-2000; 2000DE-1019173.  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIC-) EPIDENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX  
 DR WPI; 2001-602752/68.  
 PT Fragments of chemically modified genes associated with tumour suppressor  
 PT genes and oncogenes, useful in designing primers and probes for

PT analysing diseases associated with cytosine methylation state e.g.  
 PT cancer  
 XX  
 PS Claim 1; SEQ ID No 92; 27pp; English.  
 XX  
 CC The invention relates to a nucleic acid comprising a sequence of 18  
 CC bases, of a segment of chemically pretreated DNA (cp DNA) e.g. with  
 CC bisulphite, of genes associated with tumour suppression and  
 CC oncogenes having a sequence taken from 536 (actually 533 since  
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences  
 CC (SS) and sequences complementary to (SS). The nucleic acid may be a  
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may  
 CC form part of a set of probes for detecting the cytosine methylation state  
 CC and/or single nucleotide polymorphisms and also to be used in an  
 CC array for analysing diseases associated with CpG dinucleotides e.g.  
 CC cancers and tumours. The probes can also be used in a method for  
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis  
 CC and/or therapy of existing diseases or the predisposition to specific  
 CC diseases, by analysing cytosine methylations. The parameters may be  
 CC compared to another set of genetic and/or epigenetic parameters, the  
 CC differences serving as basis for diagnosis and/or prognosis events which  
 CC are disadvantageous to patients. The present sequence is one of the  
 CC 533 genomic sequences derived from tumour suppressor genes and  
 CC oncogenes. Sequences with even numbered Seq ID numbers are the  
 CC complementary sequence of the corresponding odd numbered sequence (e.g.  
 CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence  
 CC is missing).  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 6169 BP; 1840 A; 112 C; 1192 G; 3025 T; 0 other;  
 Query Match 10.9%; Score 34.2; DB 22; Length 6169;  
 Best Local Similarity 54.3%; Pred. No. 16;  
 Matches 69; Conservative 0; Mismatches 58; Indels 0; Gaps 0;  
 QY 187 atgattgtcacacaataaactgaggttattggaacatcaagtagaagaatacaacat 246  
 Db 2251 AATCATTTCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2292  
 QY 247 tccatcccttacagagatcattactgcaactcagcagataattgtcatgtgatactc 306  
 Db 2291 ACATTTCCCTTACAAATATATATATTAATAACTAAAAATTTTATCTATATTTTAAAC 2232  
 QY 307 tactat 313  
 Db 2231 TTTTAAT 2225  
 RESULT 13  
 ABL3323/C  
 ID ABL3323 standard; DNA; 9741 BP.  
 XX  
 AC ABL3323;  
 XX  
 DT 26-MAR-2002 (first entry)  
 DE Human immune system associated gene SEQ ID NO: 1296.  
 XX  
 PR Human immune system associated gene SEQ ID NO: 1296.  
 XX  
 KW Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; antihaemic; cytostatic; nootropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antineumatic; antiarthritic; antidiabetic; antipsoptic;  
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.  
 XX  
 OS Homo sapiens.  
 XX

PR 30-JUN-2000; 2000DEF-1032529.  
PR 01-SEP-2000; 2000DEF-1043826.  
PA (EPiG-) EPIGENOMICS AG.  
XX  
XX  
PI Olek A. Piepenbrock C, Berlin K;  
XX  
DR WPI; 2002-130909/17.  
XX  
PT Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation  
XX  
PS Claim 1; SEQ ID NO 1676; 32pp + Sequence Listing; German.  
XX  
CC The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC Rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.  
XX  
XX Sequence 18624 BP; 5682 A; 113 C; 3466 G; 9361 T; 2 other;

Query Match	10.9%	Score 34.2	DB 24	Length 18624
Best Local Similarity	47.9%	Pred. No. 20		
Matches 128	Conservative	0	Mismatches 138	Indels 1
QY	48	aatgaaatcaagtttccaccacacttaagaatatactcttagagatcctaagctccctta	107	
Db	3411	AAAAAATATATATTCGCCAAAAATTAATATATCTTTAAATATATAAATCCATCTTAA	3352	
QY	108	ggggacatacaagtcagtggtggtgccttggtagtcccaactataltcaagtagt	167	
Db	3351	ACCTATTTACTACACCAATTTCTTTTCCTTTTAAAAAATAAAAAATTTTAAATTACAT	3292	
QY	168	a-tgactacaattttggaanaatagatgtgcacaataaactgagttatggaacac	226	
Db	3291	ACTATTTTACATTTTAACTAACTAAATTAATTAACCAATTTTACCCCATATATTAAACGC	3232	
QY	227	agtagaaggaanaataacatccatcccttaacggaatacatcttaactgcaactcagat	286	
Db	3231	AATTTTAAACTACGTTTATTTCTTACATTATTAATTAACATTATATTAACCTTCAATT	3172	
QY	287	aatttgcatcgtgatactactaat	313	
Db	3171	ATTTATACAAATTAATAAATAAATAAATAAT	3145	

	RESULT	15
AAQ13727	AAQ13727 standard; DNA; 3088 BP.	
ID		
XX		
AC	AAQ13727;	
XX		
DT	05-DEC-1991 (first entry)	
XX		
DE	Plasmodium falciparum sporozoite antigen N-terminal	
XX		
KW	Vaccine; NXY gene; malaria; antibodies; ss.	
XX		
OS	Plasmodium falciparum.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	948..3086
FT		/tag=a
FT		/label=sporozoite-antigen
XX		/note="N-terminal"
EP447956-A.		

```

XX 25-SEP-1991.
PD
XX
XX 14-MAR-1991; 91EP-0103920.
PF
XX 23-MAR-1990; 90CH-0000970.
PR
XX (HOFF ) HOFFMANN-LA ROCHE AG.
PA
XX
XX Certa U, Guttinger M, Matile H;
PI
XX WPI: 1991-28289/39.
DR
XX P-PSDB: AARI3991.
XX
XX New antimalarial polypeptide(s) - corresp. to specific epitope(s)
PT of the Plasmodium falciparum sporozoite antigen, and are useful
PR as vaccines or to prepare antigens
XX
XX PS Disclosure: Fig 2; 36pp; German.
XX
XX The sequence is useful in the prodn. of polypeptides corresp. in
CC at least one specific epitope with the P. falciparum sporozoite
CC antigen N-terminal. These polypeptides are useful in the prodn.
CC of antimalarial vaccines.
CC The gene, designated, NXY, was isolated from a genebank produced
CC from P. falciparum cells (K1 isolate) screened with antiserum
CC obtained by immunising rabbits with P. falciparum isolate NF54.
CC See also AAQ13728.
XX
XX SQ Sequence 3088 BP; 1375 A; 273 C; 575 G; 865 T; 0 other;

```

```

Query Match          10.8%; Score 34; DB 12; Length 3088;
Best Local Similarity 61.1%; Pred. No. 15;
Matches 55; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
QY 154 atattcaagtaggtactacaatttgaataatagattgtcacacaataaactgagtt 213
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 426 ataataaataacataagaaataactaataatgataatacatgtgaacaataatcgcgaat 485
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 214 ttatggaacatcagtagagaagaatacaa 243
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 486 ctaggggcacatctgtacaaagacgcaaaa 515
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: September 16, 2002, 03:05:06  
Job time: 7984 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 03:05:06 : Search time 390.12 Seconds  
(without alignments)  
3938.887 Million cell updates/sec

Title: US-09-700-770-5

Perfect score: 895

Sequence: 1 ctaactgttactgacagc.....aataatattttaaatgac 895

Scoring table:

IDENTITY\_NNC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

N\_Geneseq\_032802:\*

1: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	895	100.0	895	21	AAZ29726 Human lung specific
2	895	100.0	895	22	AAH77949 Nucleotide sequenc
3	895	100.0	1171	22	AAH77951 Nucleotide sequenc
4	858	95.9	870	21	AAZ98202 Human signal pepti
5	856.4	95.7	1148	22	ABA09257 Human TGC-440 homo
6	852	95.2	870	21	AAZ65001 Membrane-bound pro
7	852	95.2	870	22	AAZ45994 Human DNA encoding
8	852	95.2	870	22	AAF92070 Human PRO842 CDNA
9	852	95.2	870	22	AAFA4147 Human PRO842 (UNO4

10	699	78.1	849	21	AAA08343 Human TGC-440 secr
11	630	70.4	654	21	AAZ59829 Human secreted pro
12	525.8	58.7	533	22	AAZ68195 Human lung tumour
13	445.2	49.7	511	19	AAV68618 Human secreted pro
14	362.4	40.5	365	20	AAZ40454 Human secreted pro
15	357	39.9	357	21	AAA08344 Human TGC-440 secr
16	312	34.9	312	22	AAZ68125 Human lung tumour
17	291	32.5	291	21	AAA08345 Mature human TGC-4
18	262.2	29.3	764	21	AAA08349 Mouse TGC-440 secr
19	245.6	27.4	357	21	AAA08350 Mouse TGC-440 secr
20	241.8	27.0	959	21	AAA08346 Rat TGC-440 secret
21	228.4	25.5	235	20	AAZ40453 Human secreted pro
22	228	25.5	357	21	AAA08347 Rat TGC-440 secret
23	200.4	22.4	291	21	AAA08351 Mature mouse TGC-4
24	189.2	21.1	291	21	AAA08348 Mature rat TGC-440
25	189	21.1	195	16	AAZ25820 Human gene signatu
26	73.4	8.2	97	22	AAZ07713 Cervical cancer pr
27	45.8	5.1	518	22	AAH70020 Human cervical can
28	44.6	5.0	51	21	AAA08356 Human TGC-440 secr
29	42.8	4.8	456	22	AAH71445 Human cervical can
30	39	4.4	775	22	AAK71873 Human immune/haema
31	39	4.4	775	22	AAK71874 Human immune/haema
32	38.6	4.3	356	22	AAK58069 Human immune/haema
33	36.4	4.1	704	22	AAZ13204 Human breast cance
34	36	4.0	48452	23	ABL07108 Drosophila melanog
35	35.8	4.0	47	21	AAA08355 Human TGC-440 secr
36	35.6	4.0	632	22	AAH31453 Human olfactory re
37	35.6	4.0	6792	22	AAK52133 Human polynucleoti
38	35.6	4.0	6815	22	AAH90035 Human bone marrow
39	35.6	4.0	7317	22	AAH89922 Human immune/haema
40	35.6	4.0	51469	22	AAK69322 Human immune/haema
41	35.6	4.0	51469	22	AAK70270 Human immune/haema
42	35.6	4.0	51469	22	AAK78813 Human immune/haema
43	35.4	4.0	2484	23	AAZ85729 DNA encoding novel
44	35.4	4.0	7642	24	ABL35116 Human immune syste
45	35	3.9	1610	22	AAI93220 Human polynucleoti

#### ALIGNMENTS

RESULT 1

AAZ29726 standard; DNA: 895 BP.

AAZ29726:

27-MAR-2000 (first entry)

Human lung specific gene-3.

Lung Specific Gene; LSG; human; diagnostic marker;

prognosticate; lung cancer; diagnosis; ds.

Homo sapiens.

WO9960160-A1.

25-NOV-1999.

12-MAY-1999; 99MO-US10344.

21-MAY-1998; 98US-0086212.

(DIAD-) DIADEXUS LLC.

Yang F, Macina RA, Sun Y;

WPI, 2000-116320/10.

A new method for diagnosing, monitoring and staging lung cancer

Example 1; Pages 35-36; 40pp; English.

XX The present sequence is a lung specific gene (LSG) from human  
CC clone ID 126263. The LSG has high level of tissue specificity for lungs  
CC and is overexpressed in cancerous tissues. The sequence serves as a  
CC diagnostic marker for detecting, monitoring, staging and prognosticating  
CC lung cancer. The diagnosis involves comparing levels of LSG in samples  
XX obtained from patient and normal control.

SQ Sequence 895 BP; 245 A; 250 C; 183 G; 217 T; 0 other;

Query Match 100.0%; Score 895; DB 21; Length 895;  
Best Local Similarity 100.0%; Pred. No. 2e-270;  
Matches 895; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctaactgttactgaagcaagagcgtacccctcctcgttctcgcctccaatgga 60  
DB 1 ctaactgttactgaagcaagagcgtacccctcctcgttctcgcctccaatgga 60  
QY 61 acgctgacctggactaaagacacacacagcgtgatacctgacctgatacc 120  
DB 61 acgctgacctggactaaagacacacacagcgtgatacctgacctgatacc 120  
QY 121 caggagatcagagcctccagcaggaagccttcattatatttccaacttaacgt 180  
DB 121 caggagatcagagcctccagcaggaagccttcattatatttccaacttaacgt 180  
QY 181 gcaacgacagtgtgataaagatctatctctccctccctcgttgcctcaatg 240  
DB 181 gcaacgacagtgtgataaagatctatctctccctccctcgttgcctcaatg 240  
QY 241 ctgattcattcgtctctcagcagcctgaatccaggggtcgcagagggcagagga 300  
DB 241 ctgattcattcgtctctcagcagcctgaatccaggggtcgcagagggcagagga 300  
QY 301 ggcagagcttctagagatgctccaggaagcgcccaagaatgtgagtcagaatgtg 360  
DB 301 ggcagagcttctagagatgctccaggaagcgcccaagaatgtgagtcagaatgtg 360  
QY 361 ttctctgagagcccgagagaagaattcatgacagtgtctggtgcgcaagaagcagtcg 420  
DB 361 ttctctgagagcccgagagaagaattcatgacagtgtctggtgcgcaagaagcagtcg 420  
QY 421 ccctgtatcatcattcaagggaatgtgaagaacaagacacaaaggcaccagaag 480  
DB 421 ccctgtatcatcattcaagggaatgtgaagaacaagacacaaaggcaccagaag 480  
QY 481 ccaacaagacatccagagcctgcagacaatttccaacaatgtcagctaaagaagctt 540  
DB 481 ccaacaagacatccagagcctgcagacaatttccaacaatgtcagctaaagaagctt 540  
QY 541 gcttgccttctgtagagcctcgtgagcgccacttccaataaacttcagccaaga 600  
DB 541 gcttgccttctgtagagcctcgtgagcgccacttccaataaacttcagccaaga 600  
QY 601 agacagtgagcacacccacacagacactctctccacactaactcccaactgaccc 660  
DB 601 agacagtgagcacacccacacagacactctctccacactaactcccaactgaccc 660  
QY 661 acccctaataatccagtgctcctcaaaaagcagtgtttccaagatcaattttgtgtg 720  
DB 661 acccctaataatccagtgctcctcaaaaagcagtgtttccaagatcaattttgtgtg 720  
QY 721 cttctctcattgtctctctctcgtcagctcctgagcctccctaccagagct 780  
DB 721 cttctctcattgtctctctctcgtcagctcctgagcctccctaccagagct 780  
QY 781 taagcttaattacctaagaatccaggaacgttagcttccctagctagtgtcaattaac 840  
DB 781 taagcttaattacctaagaatccaggaacgttagcttccctagctagtgtcaattaac 840  
QY 841 cttaaatgcaatcaggaagtagcaaacgaagtcataataatttttaaatgtc 895  
DB 841 cttaaatgcaatcaggaagtagcaaacgaagtcataataatttttaaatgtc 895

DB 841 cttaaatgcaatcaggaagtagcaaacgaagtcataataatttttaaatgtc 895

RESULT 2

AAH77949

ID AAH77949 standard; DNA; 895 BP.

AAH77949;

13-NOV-2001 (first entry)

Nucleotide sequence of a human Lng104 polypeptide.

Human; lung cancer specific gene; LSG; Lng104; lung cancer; ss.

Homo sapiens.

Key Location/Qualifiers

FT CDS /tag= a /product= "Lng104"

W0200161055-A2.

23-AUG-2001.

20-FEB-2001; 2001MO-US05674.

17-FEB-2000; 2000US-0183188.

(DIAD-) DIADEXUS INC.

Chen S, Sun Y, Macina RA;

WPI; 2001-529917/58.

P-PSDB; AAG63977.

New lung cancer specific gene for the treatment and diagnosis of lung

cancer -

Claim 1; Page 113; 119pp; English.

CC The present sequence represents a human lung cancer specific gene  
CC (LSG), and encodes a polypeptide designated Lng104. LSGs are useful  
CC in the treatment and diagnosis of lung cancer. The treatment of lung  
CC cancer comprises the administration of a molecule which down regulates  
CC the expression of an LSG. An immune response can be mounted against a  
CC target cell expressing an LSG. Identification of potential therapeutic  
CC agents for use in imaging and treating lung cancer which comprises  
CC screening molecules for an ability to bind to or decrease expression  
CC of an LSG relative to LSG in the absence of the agent where the ability  
CC of a molecule to bind to the LSG or decrease expression of the LSG is  
CC indicative of the molecule being useful in imaging and treating lung  
CC cancer.

SQ Sequence 895 BP; 245 A; 250 C; 183 G; 217 T; 0 other;

Query Match 100.0%; Score 895; DB 22; Length 895;  
Best Local Similarity 100.0%; Pred. No. 2e-270;  
Matches 895; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctaactgttactgaagcaagagcgtacccctcctcgttctcgcctccaatgga 60  
DB 1 ctaactgttactgaagcaagagcgtacccctcctcgttctcgcctccaatgga 60  
QY 61 acgctgacctggactaaagacacacacagcgtgatacctgacctgatacc 120  
DB 61 acgctgacctggactaaagacacacacagcgtgatacctgacctgatacc 120  
QY 121 caggagatcagagcctccagcaggaagccttcattatatttccaacttaacgt 180  
DB 121 caggagatcagagcctccagcaggaagccttcattatatttccaacttaacgt 180



```
OY 181 gcaaccagatgtgcagtgaagaattcctaattctctccctcctcctgtgtgtgcactaaag 240
    |||||||
Db 181 gcaaccagatgtgcagtgaagaattcctaattctctccctcctcctgtgtgtgcactaaag 240
OY 241 ctgatgtccatgtgtctctctagcagctgtaaatccagggtgtgcagagggcacagagcgca 300
    |||||||
Db 241 ctgatgtccatgtgtctctagcagctgtaaatccagggtgtgcagagggcacagagcgca 300
OY 301 ggcacagctcttaagagatgtgtccaggagcgcccaagaatgtgagtgcacaagatgtg 360
    |||||||
Db 301 ggcacagctcttaagagatgtgtccaggagcgcccaagaatgtgagtgcacaagatgtg 360
OY 361 ttccctgagagcccgagaaagaattcatgatcagtgcttgggctgtgcacaagaagcagtcg 420
    |||||||
Db 361 ttccctgagagcccgagaaagaattcatgatcagtgcttgggctgtgcacaagaagcagtcg 420
OY 421 cccctgtgacatcttcaaggcgaatgtgaagaacaacagacccaagaagggcacacagaag 480
    |||||||
Db 421 cccctgtgacatcttcaaggcgaatgtgaagaacaacagacccaagaagggcacacagaag 480
OY 481 ccaaaacaagcatctccagagcctgtccagcaattcttccaacaatgttaagctaaagcttt 540
    |||||||
Db 481 ccaaaacaagcatctccagagcctgtccagcaattcttccaacaatgttaagctaaagcttt 540
OY 541 gctctgctcttgtagagctctgagcgcccaactcttccaatlaaacatctctcagccaaga 600
    |||||||
Db 541 gctctgctcttgtagagctctgagcgcccaactcttccaatlaaacatctctcagccaaga 600
OY 601 agacagtgagcaacacacacagacactcttcttcccaactcactcccaactgtaccc 660
    |||||||
Db 601 agacagtgagcaacacacacagacactcttcttcccaactcactcccaactgtaccc 660
OY 661 acccccaatctctccagtgctctcaaaagaatgttttccaagatcatttgtttgtgtg 720
    |||||||
Db 661 acccccaatctctccagtgctctcaaaagaatgttttccaagatcatttgtttgtgtg 720
OY 721 cctctcctagtgctctctctcctcgtcagctcttaagcctgtgcctcccttaccagagct 780
    |||||||
Db 721 cctctcctagtgctctctctcctcgtcagctcttaagcctgtgtgcctcccttaccagagct 780
OY 781 taggcttaattcactgaaagaatccaggaaactgtagcttccctagtagtgcatttaac 840
    |||||||
Db 781 taggcttaattcactgaaagaatccaggaaactgtagcttccctagtagtgcatttaac 840
OY 841 cttaaatgcaatccagaaagttagcaaacaggaagtcataataatttttaaatgtc 895
    |||||||
Db 841 cttaaatgcaatccagaaagttagcaaacaggaagtcataataatttttaaatgtc 895

RESULT 3
AAH77951
ID AAH77951 standard; DNA; 1171 BP.
XX
XX
AC AAH77951;
XX
XX
DT 13-NOV-2001 (first entry)
DE Nucleotide sequence of a human Lng104 polypeptide.
XX
XX KW Human: lung cancer specific gene; LSG; Lng104; lung cancer; ss.
XX
XX OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH CDS 217..576
FT /*tag= a
FT /*product= "Lng104"
XX
XX PN MO200161055-A2.
XX
XX PD 23-AUG-2001.
XX
```

```
PF 20-FEB-2001; 2001MO-US05674.
XX
PR 17-FEB-2000; 2000US-0183188.
XX
XX (DIAD-) DIADEXUS INC.
XX
XX Chen S, Sun Y, Macina RA;
XX
XX WPI: 2001-529917/58.
XX
XX P-PSDB; AAG63977.
XX
XX New lung cancer specific gene for the treatment and diagnosis of lung
XX cancer -
XX
XX Claim 1; Page 117; 119pp; English.
XX
XX
XX The present sequence represents a human lung cancer specific gene
XX (LSG), and encodes a polypeptide designated Lng104. LSGs are useful
XX in the treatment and diagnosis of lung cancer. The treatment of lung
XX cancer comprises the administration of a molecule which down regulates
XX the expression of an LSG. An immune response can be mounted against a
XX target cell expressing an LSG. Identification of potential therapeutic
XX agents for use in imaging and treating lung cancer which comprises
XX screening molecules for an ability to bind to or decrease expression
XX of an LSG relative to LSG in the absence of the agent where the ability
XX of a molecule to bind to the LSG or decrease expression of the LSG is
XX indicative of the molecule being useful in imaging and treating lung
XX cancer.
XX
XX Sequence 1171 BP; 333 A; 320 C; 225 G; 293 T; 0 other;
XX
XX
XX Query Match 100.0%; Score 895; DB 22; Length 1171;
XX Best Local Similarity 100.0%; Pred. No. 2.3e-270; Gaps 0;
XX Matches 895; Conservative 0; Mismatches 0; Indels 0;
XX
XX
XX OY 1 ctaatctgtacgtlaacagcaagacagcgtcacactacctgttctcgcctccaatgga 60
    |||||||
Db 22 ctaatctgtacgtlaacagcaagacagcgtcacactacctgttctcgcctccaatgga 81
OY 61 acgctgctggaagactaaagacatagacacacagctgagatctcgtacctgaatccc 120
    |||||||
Db 82 acgctgctggaagactaaagacatagacacacagctgagatctcgtacctgaatccc 141
OY 121 caggagatcaggagcctccagcagaggaaccttccattatcttccaagaacttaagct 180
    |||||||
Db 142 caggagatcaggagcctccagcagaggaaccttccattatcttccaagaacttaagct 201
OY 181 gcaaccagatgtgcagtgaagaattcctaattctctccctcctcctgtgtgtgcactaatg 240
    |||||||
Db 202 gcaaccagatgtgcagtgaagaattcctaattctctccctcctcctgtgtgtgcactaatg 261
OY 241 ctgatgtccatgtgtctctctagcagctgtaaatccagggtgtgcagagggcacagagcgca 300
    |||||||
Db 262 ctgatgtccatgtgtctctctagcagctgtaaatccagggtgtgcagagggcacagagcgca 321
OY 301 ggcacagctcttaagagatgtgtccaggagcgcccaagaatgtgagtgcacaagatgtg 360
    |||||||
Db 322 ggcacagctcttaagagatgtgtccaggagcgcccaagaatgtgagtgcacaagatgtg 381
OY 361 ttccctgagagcccgagaaagaattcatgatcagtgcttgggctgtgcacaagaagcagtcg 420
    |||||||
Db 382 ttccctgagagcccgagaaagaattcatgatcagtgcttgggctgtgcacaagaagcagtcg 441
OY 421 cccctgtgacatcttcaaggcgaatgtgaagaacaacagacccaagaagggcacacagaag 480
    |||||||
Db 442 cccctgtgacatcttcaaggcgaatgtgaagaacaacagacccaagaagggcacacagaag 501
OY 481 ccaaaacaagcatctccagagcctgtccagcaattcttccaacaatgttaagctaaagcttt 540
    |||||||
Db 502 ccaaaacaagcatctccagagcctgtccagcaattcttccaacaatgttaagctaaagcttt 561
OY 541 gctctgctcttgtagagctctgagcgcccaactcttccaatlaaacatctcagccaaga 600
```

human signal peptide-containing proteins HSP-1 to HSP-134. HSPs have anticancer, anti-inflammatory, antimicrobial, neurotrophic, hepatotropic, neuroprotective, cardiovascular and antistatic activities, and can be used in gene therapy. HSPs can be used to treat or prevent disorders associated with decreased activity or function of HSP. Antagonists of HSP are used to treat or prevent disorders associated with increased activity or function of HSP. Such diseases include cell proliferation (including cancer), inflammation, cardiovascular, neurological, reproductive or developmental disorders, (e.g. arteriosclerosis, cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia, asthma, Crohn's disease, microbial or other infections, congestive or ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSP nucleic acids can be used for the recombinant production of HSP, for detecting HSP in standard hybridisation and amplification assays (for diagnosis and monitoring), in gene therapy, as antisense, triplex-forming or ribozyme therapeutics, for detecting related sequences or genetic variations, and for chromosomal mapping. HSP are also used to raise specific antibodies (Ab) and to screen for agonists and antagonists (potential therapeutic agents). Ab are used to diagnose, or monitor, HSP-related diseases (in usual immunoassays), as therapeutic antagonists, in competitive drug screens, and for purification of HSP from natural sources.

Sequence 870 BP; 244 A; 240 C; 177 G; 209 T; 0 other;

Query Match	95.9%;	Score 858;	DB 21;	Length 870;
Best Local Similarity	100.0%;	Pred. No. 8.2e-259;		
Matches 858; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0

QY 38 ccgtttctccgcccctcaaaatgggaaacgttgccttgcgtgactaaagcattacccaagaagctg 97  
Db 1 ccgtgtctcgcctcaaatcgggaacgtctgccttgcgtgactaaagcattacccaagaagctg 60  
QY 98 agatctcgtacccgtgactatcccccaaggatatacgaagcctccacgaaggaaacttcacata 157  
Db 61 agatctcgtacccgtgactatcccccaaggatatacgaagcctccacgaaggaaacttcacata 120  
QY 158 tattcttcaagaacacttacaagctgtgacccgcagacagtttgcgtatgaagttcttactctctccc 217  
Db 121 tattcttcaagaacacttacaagctgtgacccgcagacagtttgcgtatgaagttcttactctctccc 180

QY 218 tcctcctgttgcgcactaatgctgatgtccatggtctctctaagcagcctaataatccaaaga 277

Db 181 tcctcctgttgcgcactaatgtctgatgtlccatgtlctctagcagcctqaatccaagg 240

QY 278 tcgccagagcaccagggaccgagggccaggtctctagagatgctcccaqqaagcagcc 337

Db 241 tcgccagagccacagggaccgagggccaggtctctagagatgctccagqaagcgc 300

QY 338 aagaatgtgagtgcagaagattggtctcctgagagcccgagaagaataatcatgacattt 397

D5 301 aaagaatgtgagtcacaaagattgtgttcctgagagcccgagagaagaatattcatgacagtgt 360

QY 398 ctgggctgccaagaagcagtgccctgtgatcatttcaaggcgaatgtgaagaacaa 457

361 ctgggctgccaaagagcagtgccctgtgatcatttcaagygcaatgtgagaacaacaa 420

458 gacacccaagycaccacagaagaagccaacaagcatccagagcctgcagcaattctca 517

421 gacacccaagycaccacagaagccaacagcatcagagcctgcagcatctca 480

518 aaacatgtcagctaagaagcttctctctgccttctgtagagctctgagcgccactctc 577

481 aaacaaagtcagctaagaagcttctgtctctgccttctgtaggaagctctgagcgcccaactctc 540

5/6 caatlaacatttctcagccaagaacagttagcacaccctaccgaacactcttctctcc 637

341 caatlcaaacattctcagccagaagacagtgtgacacacctaaccagacactcttctctcc 600

536 caccclacclctcccaclgtaccacccctaataatcatctcagtgctcctcaaaaagcatgtt 697

Db 601 cactcactctccacgtctaccaccctaatactccagtgctctcaaaaagcatgt 660  
Qy 698 ttcaagaatcatgtttgtgtctctctctagtgctctctctctcgtcagcttagcc 757  
Db 661 ttcaagaatcatgtttgtgtctctctctctctctctctctctctctctcgtcagcttagcc 720  
Qy 758 tgtgcccctccctaccagagcttaggttaattaccctgaaagatccaggaactgtag 817  
Db 721 tgtgcccctccctaccagagcttaggttaattaccctgaaagatccaggaactgtag 780  
Qy 818 ctctcctagctagtgctcatttaaccttaaatgcacaggaagtgcacaacagaagtc 877  
Db 761 ctctcctagctagtgctcatttaaccttaaatgcacaggaagtgcacaacagaagtc 840  
Qy 878 taatatttttaaatgtc 895  
Db 841 taaatatttttaaatgtc 858

RESULT 5  
ABA09257  
ID ABA09257 standard; cDNA: 1148 BP.  
AC ABA09257;  
XX  
DT 11-JAN-2002 (first entry)  
XX  
XX Human TGC-440 homologue-encoding cDNA, SEQ ID NO:1033.  
DE  
XX Human; cytokine; cell proliferation; cell differentiation; growth factor;  
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;  
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;  
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
KW chronic inflammatory condition; proliferative retinopathy;  
KW bone disorders; coronary heart disease; arterial ischaemia;  
KW atherosclerosis; osteoporosis; vascular growth disorder;  
KW tissue regeneration; wound healing; infection; immune disorder;  
KW cell culture; drug screening; gene therapy; antiinflammatory;  
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;  
KW cytosolic; osteopathic; vasotropic; cardiant; virucide; antibacterial;  
KW antifungal; vulnery; antitumor; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157188-A2.  
XX  
XX 09-AUG-2001.  
PD  
XX 05-FEB-2001; 2001WO-US03800.  
PF  
XX 03-FEB-2000; 2000US-0496914.  
PR 27-APR-2000; 2000US-0560875.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX Tang YT, Liu C, Drmanac RT;  
PI  
XX WPI: 2001-457740/49.  
DR P-PSDB: ABB12013.  
XX  
XX Human proteins and DNA encoding sequences useful for preventing,  
PT treating or ameliorating a medical condition in a mammalian subject  
PT e.g. arthritis and cancer -  
XX  
XX Claim 1; Page 867-868; 1963pp; English.  
XX  
XX Sequences ABA10981-AB812330 represent 1350 novel human polypeptides, and  
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
CC invention also relates to vectors and recombinant host cells comprising a  
CC nucleotide of the invention, methods of producing the novel polypeptides,  
CC antibodies against the polypeptides, methods of detecting the nucleotides  
CC or polypeptides in a sample, and methods of identifying compounds which

CC bind to polypeptides of the invention. Although novel, many of the  
CC polypeptides of the invention have homology to known proteins, thereby  
CC giving an insight into their probable biological activities, and hence  
CC potential therapeutic applications. The polypeptides of the invention may  
CC have various activities, including cytokine, cell proliferation or cell  
CC differentiation activities; stem cell growth factor activity;  
CC haematopoiesis regulatory activity; tissue growth activity;  
CC immunomodulatory activity; activin- or inhibin-related activities;  
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or  
CC thrombolytic activities; receptor or ligand activities; or may be  
CC involved in oncogenesis, cancer cell proliferation or metastasis.  
CC Depending on their biological activities, polypeptides and nucleotides of  
CC the invention are useful for preventing, treating or ameliorating medical  
CC conditions, e.g., by protein or gene therapy. Such conditions include  
CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
CC vascular growth. Polypeptides involved with tissue regeneration and  
CC repair (or nucleic acids encoding them) may be used to promote wound  
CC healing (e.g., of burns, incisions and ulcers), while those with  
CC immunomodulatory activities may be used in the treatment of viral,  
CC bacterial and fungal infections in addition to immune disorders.  
CC Polypeptides with growth factor activity may be used in cell cultures to  
CC promote cell growth. For example, such polypeptides may be used to  
CC manipulate stem cells in culture to give rise to neuroepithelial cells  
CC that can be used to augment or replace cells damaged by illness,  
CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
CC may also be used in the diagnosis of the above conditions, and in drug  
CC screening techniques. The present sequence represents a cDNA encoding a  
CC novel human polypeptide of the invention.  
SQ Sequence 1148 BP: 328 A; 309 C; 222 G; 289 T; 0 other;

Query Match 95.7%; Score 856.4; DB 22; Length 1148;  
Best Local Similarity 99.9%; Pred. NO. 3e-258;  
Matches 857; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 38 cctgtctcgccctcaaatgtagacgtgctgctggactaaagcatagaccacagctgt 97  
Db 23 cctgtctcgccctcaaatgtagacgtgctgctggactaaagcatagaccacagctgt 82  
Qy 98 agtaccgagcttgatatacccaaggatcagagccctccagcagggaaccttcacata 157  
Db 83 agtaccgagcttgatatacccaaggatcagagccctccagcagggaaccttcacata 142  
Qy 158 tatcttcaagcaacttaacagctgcacgcagctgtgcagtgaaagtctcaatctctccc 217  
Db 143 tatcttcaagcaacttaacagctgcacgcagctgtgcagtgaaagtctcaatctctccc 202  
Qy 218 tctctcgtgtgtcgcactaatgtgtatgtccatgtctctagcagcctgtaatccaagg 277  
Db 203 tctctcgtgtgtcgcactaatgtgtatgtccatgtctctagcagcctgtaatccaagg 262  
Qy 278 tgcgcaaggcgcacagggacccagggcctctcagagataggtctccaggaagcgggcc 337  
Db 263 tgcgcaaggcgcacagggacccagggcctctcagagataggtctccaggaagcgggcc 322  
Qy 338 aagaatgtgagtcacaaagtgtctcttagagcccccagaaagaataatcagaaagtgt 397  
Db 323 aagaatgtgagtcacaaagtgtctcttagagagcccccagaaagaataatcagaaagtgt 382  
Qy 398 ctgggctgcacaaagacagtgccctgtgatcatcttcaagggcaatgtgaagaacaa 457  
Db 383 ctgggctgcacaaagacagtgccctgtgatcatcttcaagggcaatgtgaagaacaa 442  
Qy 458 gacaccaagagcaccacaggaagccaaacaagcatctcagagcctgcgaacattctca 517  
Db 443 gacaccaagagcaccacaggaagccaaacaagcatctcagagcctgcgaacattctca 502  
Qy 518 aacaatgcagctagaagccttgctgcctttagaagctcgaagccactcttc 577



PR 04-AUG-1998; 98US-0095325.  
 PR 10-AUG-1998; 98US-0095916.  
 PR 10-AUG-1998; 98US-0095929.  
 PR 10-AUG-1998; 98US-0096012.  
 PR 11-AUG-1998; 98US-0096143.  
 PR 11-AUG-1998; 98US-0096146.  
 PR 12-AUG-1998; 98US-0096329.  
 PR 17-AUG-1998; 98US-0096757.  
 PR 17-AUG-1998; 98US-0096766.  
 PR 17-AUG-1998; 98US-0096768.  
 PR 17-AUG-1998; 98US-0096773.  
 PR 17-AUG-1998; 98US-0096791.  
 PR 17-AUG-1998; 98US-0096867.  
 PR 17-AUG-1998; 98US-0096894.  
 PR 17-AUG-1998; 98US-0096895.  
 PR 17-AUG-1998; 98US-0096897.  
 PR 18-AUG-1998; 98US-0096949.  
 PR 18-AUG-1998; 98US-0096950.  
 PR 18-AUG-1998; 98US-0096959.  
 PR 18-AUG-1998; 98US-0096960.  
 PR 18-AUG-1998; 98US-0097022.  
 PR 19-AUG-1998; 98US-0097141.  
 PR 20-AUG-1998; 98US-0097218.  
 PR 20-AUG-1998; 98US-0097661.  
 PR 26-AUG-1998; 98US-0097951.  
 PR 26-AUG-1998; 98US-0097952.  
 PR 26-AUG-1998; 98US-0097954.  
 PR 26-AUG-1998; 98US-0097955.  
 PR 26-AUG-1998; 98US-0097971.  
 PR 26-AUG-1998; 98US-0097974.  
 PR 26-AUG-1998; 98US-0097978.  
 PR 26-AUG-1998; 98US-0097979.  
 PR 26-AUG-1998; 98US-0097986.  
 PR 26-AUG-1998; 98US-0098014.  
 PR 31-AUG-1998; 98US-0098525.  
 PR 16-SEP-1998; 98US-0100634.  
 PR 12-JAN-1999; 99US-0115565.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;  
 PI Wood WJ, Yuan J;  
 XX  
 DR WPI: 2000-072883/06.  
 XX P-PSDB: AAY6668.  
 PR  
 PT Membrane-bound proteins and related nucleotide sequences -  
 PS  
 XX Claim 2; Fig 98; 822pp; English.  
 CC The invention provides membrane-bound PRO polypeptides and  
 CC polynucleotides encoding them. The PRO sequences of the invention were  
 CC identified based on extracellular domain homology screening. The PRO  
 CC sequences have homology with proteins including LDL receptors, TIE  
 CC ligands and various enzymes. The membrane-bound proteins and receptor  
 CC molecules are useful as pharmaceutical and diagnostic agents. Receptor  
 CC immunoadhesins, for instance, can be used as therapeutic agents to block  
 CC receptor-ligand interactions. The membrane-bound proteins can also be  
 CC employed for screening of potential peptide or small molecule inhibitors  
 CC of the relevant receptor/ligand interaction. The PRO encoding sequences  
 CC are useful as hybridization probes, in chromosome and gene mapping and in  
 CC the generation of antisense RNA and DNA. PRO nucleic acid sequences  
 CC will also be useful for the preparation of PRO polypeptides, especially  
 CC by recombinant techniques.  
 XX  
 SQ Sequence 870 BP; 251 A; 237 C; 176 G; 206 T; 0 other;

Query Match 95.2%; Score 852; DB 21; Length 870;  
 Best Local Similarity 100.0%; Pred. No. 6,3e-257;  
 Matches 852; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 ctgcgacctcaaatggaacgcttgccctggactaaagcatagaccacagcctgaatc 103  
 DB 1 ctgcgacctcaaatggaacgcttgccctggactaaagcatagaccacagcctgaatc 60  
 QY 104 ctgacactgagatcccccaggatccagagcctccagcagggaaccttcataatct 163  
 DB 61 ctgacactgagatcccccaggatccagagcctccagcagggaaccttcataatct 120  
 QY 164 tcaagcaacttaacgtcagtcacgcagacagttgcgtatgaatcttaactctccctcc 223  
 DB 121 tcaagcaacttaacgtcagtcacgcagacagttgcgtatgaatcttaactctccctcc 180  
 QY 224 tcttgctgcacttaatgctcgtatgctcactgagcagcctgaatccagggttcgca 283  
 DB 181 tcttgctgcacttaatgctcgtatgctcactgagcagcctgaatccagggttcgca 240  
 QY 284 gaggccacagggagccgagccagcagctctcagagatgctccagagagcgcccaagaat 343  
 DB 241 gaggccacagggagccgagccagcagctctcagagatgctccagagagcgcccaagaat 300  
 QY 344 gtgagtgcacaagattggttcctcgagagcccccagagaagaatctatgacagtgtcggc 403  
 DB 301 gtgagtgcacaagattggttcctcgagagcccccagagaagaatctatgacagtgtcggc 360  
 QY 404 tgcacaagaagcagtgccctgtgatcatctcaaggcgcaatgtgaagaacaagacacc 463  
 DB 361 tgcacaagaagcagtgccctgtgatcatctcaaggcgcaatgtgaagaacaagacacc 420  
 QY 464 aaagycaccacagaaaagccaaacaagcaltccagagcctgcagcaattctcnaacaat 523  
 DB 421 aaagycaccacagaaaagccaaacaagcaltccagagcctgcagcaattctcnaacaat 480  
 QY 524 gtcaagtaagaagcttgctcgtccttgtaagagctcgcagcgcacactctccaatta 583  
 DB 481 gtcaagtaagaagcttgctcgtccttgtaagagctcgcagcgcacactctccaatta 540  
 QY 584 aacattctcagccaagaagaacagtgagcacactaccagacactctctccaccctc 643  
 DB 541 aacattctcagccaagaagaacagtgagcacactaccagacactctctccaccctc 600  
 QY 644 actctccacgttacccacccttaaatcatctcaagtgtctcnaaagaagcttttcaaa 703  
 DB 601 actctccacgttacccacccttaaatcatctcaagtgtctcnaaagaagcttttcaaa 660  
 QY 704 gatcatcttggttggtgcctctctagtgctctctctctctctctctctctctctct 763  
 DB 661 gatcatcttggttggtgcctctctagtgctctctctctctctctctctctctctct 720  
 QY 764 ctccaccttaacagcgttagcttaattacctggaagatccaggaactgtagctctc 823  
 DB 721 ctccaccttaacagcgttagcttaattacctggaagatccaggaactgtagctctc 780  
 QY 824 agctagtgcatcttaacctaataatgcaatcagagaagtgcacaagaagtcaataata 883  
 DB 781 agctagtgcatcttaacctaataatgcaatcagagaagtgcacaagaagtcaataata 840  
 QY 884 ttttaaatgctc 895  
 DB 841 ttttaaatgctc 852  
 XX  
 AC AAS45994;  
 XX  
 DT 18-DEC-2001 (first entry)  
 XX  
 DE Human DNA encoding PRO polypeptide sequence #70.  
 XX PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep; ss;  
 KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;

KM blood; chondrocyte cell; cell proliferation; cell differentiation; colon;  
 KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder;  
 XX PCR primer.  
 OS Homo sapiens.  
 XX  
 PN M0200168848-A2.  
 XX  
 XX 20-SEP-2001.  
 PD  
 XX  
 PF 28-FEB-2001; 2001MO-US06520.  
 XX  
 PR 01-MAR-2000; 2000MO-US05601.  
 PR 02-MAR-2000; 2000MO-US05841.  
 PR 03-MAR-2000; 2000US-187202P.  
 PR 06-MAR-2000; 2000US-186968P.  
 PR 14-MAR-2000; 2000US-189320P.  
 PR 14-MAR-2000; 2000US-189328P.  
 PR 15-MAR-2000; 2000MO-US06884.  
 PR 21-MAR-2000; 2000US-190828P.  
 PR 21-MAR-2000; 2000US-191007P.  
 PR 21-MAR-2000; 2000US-191048P.  
 PR 21-MAR-2000; 2000US-191314P.  
 PR 28-MAR-2000; 2000US-192655P.  
 PR 29-MAR-2000; 2000US-193032P.  
 PR 29-MAR-2000; 2000US-193053P.  
 PR 30-MAR-2000; 2000MO-US08439.  
 PR 04-APR-2000; 2000US-194449P.  
 PR 11-APR-2000; 2000US-194647P.  
 PR 11-APR-2000; 2000US-195975P.  
 PR 11-APR-2000; 2000US-196000P.  
 PR 11-APR-2000; 2000US-196187P.  
 PR 11-APR-2000; 2000US-196690P.  
 PR 11-APR-2000; 2000US-196820P.  
 PR 18-APR-2000; 2000US-198121P.  
 PR 18-APR-2000; 2000US-198585P.  
 PR 25-APR-2000; 2000US-199397P.  
 PR 25-APR-2000; 2000US-199550P.  
 PR 25-APR-2000; 2000US-199554P.  
 PR 03-MAY-2000; 2000US-201516P.  
 PR 17-MAY-2000; 2000MO-US13705.  
 PR 22-MAY-2000; 2000MO-US14042.  
 PR 30-MAY-2000; 2000MO-US14942.  
 PR 02-JUN-2000; 2000MO-US15264.  
 PR 05-JUN-2000; 2000US-209832P.  
 PR 28-JUL-2000; 2000MO-US20710.  
 PR 22-AUG-2000; 2000US-0644848.  
 PR 24-AUG-2000; 2000MO-US23328.  
 PR 08-NOV-2000; 2000MO-US30952.  
 PR 01-DEC-2000; 2000MO-US32678.  
 PR 20-DEC-2000; 2000MO-US34956.  
 XX  
 XX (GETH ) GENENTECH INC.  
 PA  
 PI Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;  
 PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;  
 XX  
 XX WPI; 2001-602746/68.  
 DR  
 DR P-PSDB; AAU29093.  
 XX  
 XX Novel nucleic acids encoding PRO polypeptides, used to diagnose the  
 PT presence of tumours, such as prostate and breast tumours, in mammals and  
 PT to screen for modulators of the compounds -  
 XX  
 XX Claim 2; Fig 139; 774pp; English.  
 PS  
 XX Sequences AAS45925-AAS4621 represent DNA molecules encoding and PCR  
 CC primers for PRO polypeptides of the invention. The sequences of the  
 CC invention can be used to detect the presence of a tumour in a mammal by  
 CC comparing the level of expression of a PRO polypeptide in a test sample  
 CC of cells from the animal and a control sample of normal cells, whereby a  
 CC higher level of expression in the test sample indicates the presence of a  
 CC tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep,

CC pigs, goats and rabbits but are preferably human. The polypeptides can be  
 CC used to stimulate tumour necrosis factor (TNF) alpha release from human  
 CC blood, when contracted with it. A specific polypeptide can be used to  
 CC stimulate the proliferation or differentiation of chondrocyte cells. The  
 CC PRO proteins can be used to determine the presence of tumours and also  
 CC susceptibility to tumour development, particularly adrenal, lung, colon,  
 CC breast, prostate, rectal, cervical, or liver tumours. In mammalian  
 CC subjects. The oligonucleotide probes specific for the PRO nucleic acids  
 CC can be used for genetic analysis of individuals with genetic disorders.  
 XX  
 XX Sequence 870 BP; 251 A; 237 C; 176 G; 206 T; 0 other;  
 SQ  
 Query Match 95.2%; Score 852; DB 22; Length 870;  
 Best Local Similarity 100.0%; Pred. No. 6.3e-257;  
 Matches 852; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 44 ctgcgcctcaaatggaacgtctgctcgtggaactaaagcatagaccacagctgagatc 103  
 Db 1 ctgcgcctcaaatggaacgtctgctcgtggaactaaagcatagaccacagctgagatc 60  
 QY 104 ctgacctgagtcaccccccagagatcaagagccctccagcagggaaccttcattattct 163  
 Db 61 ctgacctgagtcaccccccagagatcaagagccctccagcagggaaccttcattattct 120  
 QY 164 tcaagcaacttaacgctgcacgcagcaagcttgagatgaagttcaatctctccctcc 223  
 Db 121 tcaagcaacttaacgctgcacgcagcaagcttgagatgaagttcaatctctccctcc 180  
 QY 224 tgtgtcgtccactaatgctatgtatcatalgtctctcgtcagcgtcgtatccagggtgc 283  
 Db 181 tgtgtcgtccactaatgctatgtatcatalgtctctcgtcagcgtcgtatccagggtgc 240  
 QY 284 gaagccaaggaacgagcagagcagctctcagagatgagtcacaggaagcgccagaat 343  
 Db 241 gaagccaaggaacgagcagagcagctctcagagatgagtcacaggaagcgccagaat 300  
 QY 344 gtgagtcaaaagtgtgttcctcgtgagagcccgagaagaattcatgacagtgtctggc 403  
 Db 301 gtgagtcaaaagtgtgttcctcgtgagagcccgagaagaattcatgacagtgtctggc 360  
 QY 404 tgcgaagaagcagtgccctctgcatcatctcaagagcaattgaaagaacacaccc 463  
 Db 361 tgcgaagaagcagtgccctctgcatcatctcaagagcaattgaaagaacacacaccc 420  
 QY 464 aaagcaccacagaaagcacaacagcatccagagcctgcagcaatttccaacaat 523  
 Db 421 aaagcaccacagaaagcacaacagcatccagagcctgcagcaatttccaacaat 480  
 QY 524 gtccagctaagaagcttctcgtccttctgtaggaagctctgagcgccacttccaat 583  
 Db 481 gtccagctaagaagcttctcgtccttctgtaggaagctctgagcgccacttccaat 540  
 QY 584 aacattctcagccaagaagaagcagtgagcacacactaccagaaactctctccacctc 643  
 Db 541 aacattctcagccaagaagaagcagtgagcacacactaccagaaactctctccacctc 600  
 QY 644 actctccactgtaccacccactaatatcatccagtgctctccaaaagcagtgtttcaa 703  
 Db 601 actctccactgtaccacccactaatatcatccagtgctctccaaaagcagtgtttcaa 660  
 QY 704 gatacttctgtctgtctctcctcctagtgctctctctcctcgtcagctgtgctc 763  
 Db 661 gatacttctgtctgtctctcctcctagtgctctctctcctcgtcagctgtgctc 720  
 QY 764 ctcccttaaccagccttagcttaattatcctgaagaattccaggaactgtagctcct 823  
 Db 721 ctcccttaaccagccttagcttaattatcctgaagaattccaggaactgtagctcct 780  
 QY 824 agctagtgatcaattcaacttaaatgacatcaggaagtagcaagaagtgcaataata 883  
 Db 781 agctagtgatcaattcaacttaaatgacatcaggaagtagcaagaagtgcaataata 840

OY 884 ttlttaaatgtc 895  
|||||  
Db 841 ttlttaaatgtc 852

RESULT 8  
AAf92070  
ID AAF92070 standard; cDNA; 870 BP.  
XX  
AC AAF92070;  
XX  
DT 15-MAY-2001 (first entry)  
XX  
DE Human PRO842 cDNA.  
XX  
KW Human; PRO protein; mapping; ss.  
XX  
OS Homo sapiens.  
XX  
PN MO200116318-A2.  
XX  
PD 08-MAR-2001.  
PF 24-AUG-2000; 2000MO-US23328.  
XX  
PR 01-SEP-1999; 99MO-US20111.  
PR 15-SEP-1999; 99MO-US21090.  
PR 07-DEC-1999; 99US-0169495.  
PR 09-DEC-1999; 99US-0170262.  
PR 11-JAN-2000; 2000US-0175481.  
PR 18-FEB-2000; 2000MO-US04341.  
PR 18-FEB-2000; 2000MO-US04342.  
PR 22-FEB-2000; 2000MO-US04414.  
PR 01-MAR-2000; 2000MO-US05601.  
PR 03-MAR-2000; 2000US-0187202.  
PR 25-APR-2000; 2000US-0199397.  
PR 22-MAY-2000; 2000MO-US14042.  
PR 05-JUN-2000; 2000US-0209832.  
XX  
PA (GENE ) GENENTECH INC.  
PI Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;  
PI Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;  
XX  
DR MPI: 2001-183260/18.  
DR P-PSDB; AAB87538.  
XX  
PT Eighty four nucleic acids encoding PRO polypeptides, useful in  
PT molecular biology, including use as hybridization probes, and in  
PT chromosome and gene mapping. -  
XX  
PS Claim 2; Fig 25; 278bp; English.  
XX  
CC The present sequence is the coding sequence for a human PRO polypeptide  
CC (secreted and transmembrane). The PRO protein, and PRO agonists, PRO  
CC antagonists or anti-PRO antibodies are useful for preparation of a  
CC medicament useful in the treatment of a condition which is responsive to  
CC the PRO protein, agonists, antagonists or anti-PRO antibodies. The PRO  
CC protein may also be employed as molecular weight markers for protein  
CC electrophoresis. The PRO coding sequence has applications in molecular  
CC biology, including use as hybridisation probes, and in chromosome and  
CC gene mapping.  
XX  
SQ Sequence 870 BP; 251 A; 237 C; 176 G; 206 T; 0 other;

Query Match 95.2%; Score 852; DB 22; Length 870;  
Best Local Similarity 100.0%; Pred. No. 6,3e-257;  
Matches 852; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 44 ctgcgcctcaaatgggaacgctgctgggactaagatagacacacagctgagtc 103  
|||||  
Db 1 ctgcgcctcaaatgggaacgctgctgggactaagatagacacacagctgagtc 60

OY 104 ctgacctgagtcattccccaggatcaaggacctccagcaggacattccattatct 163  
|||||  
Db 61 ctgacctgagtcattccccaggatcaaggacctccagcaggacattccattatct 120  
OY 164 tcaagcaacttaacagctgacccagacagcttgcatgtaaaattcattctccctctcc 223  
|||||  
Db 121 tcaagcaacttaacagctgacccagacagcttgcatgtaaaattcattctccctctcc 180  
OY 224 tgttgcctgcaactaaatgctgcatgctcattgctctctagcagcctgaaatccagggtcgca 283  
|||||  
Db 181 tgttgcctgcaactaaatgctgcatgctcattgctctctagcagcctgaaatccagggtcgca 240  
OY 284 gaggcacacaggagccagagccagcgtctcaggatagcttccaaggacggccaagaat 343  
|||||  
Db 241 gaggcacacaggagccagagccagcgtctcaggatagcttccaaggacggccaagaat 300  
OY 344 gtgagtgcaaaagattgctcctgtagagcccgagaaataatcattgacagtgctgggc 403  
|||||  
Db 301 gtgagtgcaaaagattgctcctgtagagcccgagaaataatcattgacagtgctgggc 360  
OY 404 tgcacaaagaagcagtgccctgtgatcatttcaaggcgaatgtgaaagaaacaagacacc 463  
|||||  
Db 361 tgcacaaagaagcagtgccctgtgatcatttcaaggcgaatgtgaaagaaacaagacacc 420  
OY 464 aaagcaccacacagaaagccaaacaagcattccaagcctgcagacatttccaacaat 523  
|||||  
Db 421 aaagcaccacacagaaagccaaacaagcattccaagcctgcagacatttccaacaat 480  
OY 524 gtcaagctaaagaagcttgctctgaccttgtagagctcgagagccactctccaatta 583  
|||||  
Db 481 gtcaagctaaagaagcttgctctgaccttgtagagctcgagagccactctccaatta 540  
OY 584 aacatctcagccaagaagacagtgagcacacctaccagacactcttctccacctc 643  
|||||  
Db 541 aacatctcagccaagaagacagtgagcacacctaccagacactcttctccacctc 600  
OY 644 actctcccactgtaacccaccttaaatcattccagtgctcctcaaaaagatgttttcaa 703  
|||||  
Db 601 actctcccactgtaacccaccttaaatcattccagtgctcctcaaaaagatgttttcaa 660  
OY 704 gatcatttgttgtgtcctcctctagtgctctctctctctctgcaactctagcctgccc 763  
|||||  
Db 661 gatcatttgttgtgtcctcctctagtgctctctctctctctgcaactctagcctgccc 720  
OY 764 ctccccttaacccaagcgttagcttaattacactgtaaaagattccaagaaactgtagctctc 823  
|||||  
Db 721 ctccccttaacccaagcgttagcttaattacactgtaaaagattccaagaaactgtagctctc 780  
OY 824 agctagtgcatcttaacacttaattgcaatcaggaagtagcaaacagagctcaataata 883  
|||||  
Db 781 agctagtgcatcttaacacttaattgcaatcaggaagtagcaaacagagctcaataata 840  
OY 884 ttlttaaatgtc 895  
|||||  
Db 841 ttlttaaatgtc 852

RESULT 9  
AAf44147  
ID AAF44147 standard; cDNA; 870 BP.  
XX  
AC AAF44147;  
XX  
DT 02-APR-2001 (first entry)  
XX  
DE Human PRO842 (UNQ473) nucleotide sequence SEQ ID NO:164.  
XX  
KW Human; secreted and transmembrane protein; PRO; cytosstatic;  
KW cell death; cancer; Chromosomal mapping; gene mapping; tissue typing;  
KW diagnostic assay; ss.  
XX  
OS Homo sapiens.

XX WO200073454-A1.  
XX  
XX 07-DEC-2000.  
XX  
XX 30-MAR-2000; 2000WO-US08439.  
XX  
PR 02-JUN-1999; 99WO-US12252.  
PR 23-JUN-1999; 99US-0141037.  
PR 07-JUL-1999; 99US-0143048.  
PR 20-JUL-1999; 99US-0144758.  
PR 26-JUL-1999; 99US-0145698.  
PR 28-JUL-1999; 99US-0146222.  
PR 17-AUG-1999; 99US-0149396.  
PR 15-SEP-1999; 99WO-US21090.  
PR 15-SEP-1999; 99US-0215457.  
PR 08-OCT-1999; 99WO-US18663.  
PR 30-NOV-1999; 99WO-US28313.  
PR 01-DEC-1999; 99WO-US28301.  
PR 16-DEC-1999; 99WO-US30095.  
PR 20-DEC-1999; 99WO-US30911.  
PR 05-JAN-2000; 2000WO-US00219.  
PR 06-JAN-2000; 2000WO-US00376.  
PR 11-FEB-2000; 2000WO-US03565.  
PR 18-FEB-2000; 2000WO-US04341.  
PR 22-FEB-2000; 2000WO-US04414.  
PR 24-FEB-2000; 2000WO-US04914.  
PR 24-FEB-2000; 2000WO-US05004.  
PR 02-MAR-2000; 2000WO-US05841.  
PR 15-MAR-2000; 2000WO-US06884.  
PR 20-MAR-2000; 2000WO-US07377.  
XX  
XX (GERTH ) GENENTECH INC.  
XX  
PI Ashkenazi AJ, Baker KP, Bolstein D, Desnoyers L, Eaton DL;  
PI Ferrara N, Fong S, Gerber H, Gerritsen ME, Goddard A, Godowski PJ;  
PI Grimaldi CJ, Gurney AL, Kijavini IU, Napier MA, Pan J, Paoni NF;  
PI Roy MA, Stewart TA, Tunas D, Watanabe CK, Williams PM, Wood WI;  
PI Zhang Z;  
XX  
DR WPI: 2001-032160/04.  
DR P-PSDB: AAB65191.  
XX  
XX PRO polynucleotides used to produce polypeptides used to target  
PT bioactive molecules such as toxins, radiolabels or antibodies, to  
PT specific cells, to cause targeted cell death .  
XX  
PS Claim 2; Fig 98; 935pp; English.  
XX  
XX The present invention describes human secreted and transmembrane PRO  
CC proteins. The PRO proteins have cytosolic activity. The PRO proteins  
CC can be used for targeted delivery of bioactive molecules, such as  
CC toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide  
CC sequences, and their fragments, can be used as hybridisation probes, in  
CC chromosomal and gene mapping, and in the generation of anti-sense RNA  
CC and DNA. They may also be used to produce transgenic animals which are  
CC used to develop and screen therapeutically useful reagents. The PRO  
CC nucleotide and protein sequence can be used for tissue typing and in  
CC treating cancer. Anti-PRO antibodies can be used in diagnostic assays.  
CC AAF44270 to AAF44470 represent PCR primers and hybridisation probes used  
CC in the isolation of human PRO sequences. AAF44087 to AAF44269 and  
CC AAB65134 to AAB65300 represent human PRO polynucleotide and protein  
CC sequences given in the exemplification of the present invention.  
XX  
SQ Sequence 870 BP; 251 A; 237 C; 176 G; 206 T; 0 other;

Query Match 95.2%; Score 852; DB 22; Length 870;  
Best Local Similarity 100.0%; Pred. No. 6.3e-257;  
Matches 852; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 ctgcgcctcaaatggaagcgtgctggaactaagcatagaccacgagctgagtc 103  
|||||

Db 1 ctgcgcctcaaatggaagcgtgctggaactaagcatagaccacgagctgagtc 60  
QY 104 ctgacctgagtcattccccaaggagatcaggagcctcccaaggaaccttcattatctc 163  
Db 61 ctgacctgagtcattccccaaggagatcaggagcctcccaaggaaccttcattatctc 120  
QY 164 tcaagcaacttaacgctgcagcccgagcaattgcgataagaaagtcttaattcttcctcc 223  
Db 121 tcaagcaacttaacgctgcagcccgagcaattgcgataagaaagtcttaattcttcctcc 180  
QY 224 tgttgcctcaactaatgctatgctatgctccttaagcagcctgaatccagggttcgcca 283  
Db 181 tgttgcctcaactaatgctatgctatgctccttaagcagcctgaatccagggttcgcca 240  
QY 284 gaggcccaaggagcagccgagccagcctctagagatgctccaggaagcgccgaagaat 343  
Db 241 gaggcccaaggagcagccgagccagcctctagagatgctccaggaagcgccgaagaat 300  
QY 344 gtgagtgcaagatgtgtctctgaagcccccgaagaataatcatgacagtgctcggc 403  
Db 301 gtgagtgcaagatgtgtctctgaagcccccgaagaataatcatgacagtgctcggc 360  
QY 404 tgcgaagaagcagtgccctgtgataattcgaaggcaatgtgaagaacaagacacc 463  
Db 361 tgcgaagaagcagtgccctgtgataattcgaaggcaatgtgaagaacaagacacc 420  
QY 464 aaaggccacacagaagaagccaaacagatctccagagcctgcagcaattctcaacaat 523  
Db 421 aaaggccacacagaagaagccaaacagatctccagagcctgcagcaattctcaacaat 480  
QY 524 gtcaagtaagaagcttgctctgtcctcttgaagagctcgaagccactcttcaata 583  
Db 481 gtcaagtaagaagcttgctctgtcctcttgaagagctcgaagccactcttcaata 540  
QY 584 aacattctcagccaaagaacagatgagacacactaccagacattcttccaccctc 643  
Db 541 aacattctcagccaaagaacagatgagacacactaccagacattcttccaccctc 600  
QY 644 actctccactgaccacccccaataatcattcagagctctcaaaaagatgttttcaa 703  
Db 601 actctccactgaccacccccaataatcattcagagctctcaaaaagatgttttcaa 660  
QY 704 gatcaattgttgtctctctctagtgctctctctctcgtcagcttaacgtgtgc 763  
Db 661 gatcaattgttgtctctctctagtgctctctctctcgtcagcttaacgtgtgc 720  
QY 764 ctcccttaaccaggttggttaattaccctggaagaatccaggaacatgtagctctc 823  
Db 721 ctcccttaaccaggttggttaattaccctggaagaatccaggaacatgtagctctc 780  
QY 824 agctagtgcatltaaccttaataatgcaatcagaagaatgacaaagatcaataaata 883  
Db 781 agctagtgcatltaaccttaataatgcaatcagaagaatgacaaagatcaataaata 840  
QY 884 ttlttaaatgctc 895  
Db 841 ttlttaaatgctc 852

RESULT 10  
AAA08343  
ID AAA08343 standard; DNA; 849 BP.  
XX  
XX AAA08343;  
XX  
DT 30-JUN-2000 (first entry)  
XX  
DE Human TGC-440 secretory protein nucleotide sequence.  
XX  
KW TGC-440; secretory protein; immunological disease; infectious disease;  
KW pulmonary function disorder; hepatic function disorder; nephrotropic;  
KW gastrointestinal function disorder; antiinflammatory; immunomodulatory;  
KW virucide; hepatotropic; antistimatic; antibacterial; vaccine;





PI Valenzuela D, Yuan O, Hoffman H, Hall J, Rapiejko P;  
 XX MPI: 2000-638211/61.  
 DR P-PSDB: AAB34728.  
 XX  
 PT Novel proteins and polypeptides useful for the treatment of e.g  
 PT multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,  
 PT cancer, Alzheimer's disease, Parkinson's disease, stroke, anemia and  
 PT ulcers  
 XX  
 PS Claim 92: Page 441: 493pp: English.

CC This invention relates to 59 human secreted proteins and the nucleotide  
 CC sequences encoding them. Sequences AAC59788-599846 and AAB34687-134745  
 CC represent the proteins and their encoding nucleotide sequences, and  
 CC sequences AAB34746-834771 represent fragments of the proteins. Probes  
 CC for the DNA sequences are represented by sequences AAC59847-C59596. The  
 CC proteins exhibit neuroprotective, dermatological, immunosuppressive,  
 CC antiinflammatory, antinaeemic, nootropic, antiparkinsonian,  
 CC antidiabetic, antihypertensive, antidiabetic, antidiabetic, antidiabetic,  
 CC antibacterial, virucide, and fungicide activity. The proteins and  
 CC nucleotide sequences are useful as nutritional sources or supplements  
 CC and in research. The proteins are useful for treating immune deficiency  
 CC and disorders, which may be genetic or resulting from infections,  
 CC autoimmune disorders such as multiple sclerosis, systemic lupus  
 CC erythematosus, rheumatoid arthritis, and for treating myeloid or lymphoid  
 CC cell deficiencies such as anaemias by regulating haematopoiesis. The  
 CC proteins are also useful in compositions for bone, cartilage, tendon,  
 CC ligament and/or nerve tissue growth or regeneration, for wound healing,  
 CC tissue repair and replacement and in the treatment of wounds, incisions  
 CC and ulcers. Other uses include in the treatment of central and  
 CC peripheral nervous system and neuropathies such as Alzheimer's and  
 CC Parkinson's diseases and Shy-Drager syndrome, and mechanical and  
 CC traumatic disorders, such as spinal cord disorders, head trauma and  
 CC stroke. The proteins may also be used as a contraceptive, and for  
 CC treating coagulation disorders such as haemophilias. The protein and  
 CC nucleotide sequences with cadherin activity are useful for treating  
 CC cancer. Other uses for the protein include for inhibiting the growth,  
 CC infection or function of, or killing, infectious agents such as bacteria,  
 CC virus, fungi and other parasites, for effecting bodily characteristics  
 CC such as height, weight, hair colour, effecting biorhythms or cardiac  
 CC cycles or rhythms, effecting metabolism, catabolism, anabolism,  
 CC processing, utilization, storage or elimination of dietary fat, lipid,  
 CC protein, carbohydrate, vitamins, minerals, cofactors, effecting  
 CC behavioural characteristics, providing analgesic effects and for treating  
 CC hyperproliferative disorders such as psoriasis.

XX Sequence 654 BP: 200 A; 183 C; 136 G; 135 T; 0 other:

Query Match 70.4%; Score 630; DB 21; Length 654;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-187;  
 Matches 630; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 agcatagaccacagagctgagtcctgacatgacatcccaaggagtaagagcctcc 138  
 DB 2 agcatagaccacagagctgagtcctgacatgacatcccaaggagtaagagcctcc 61  
 QY 139 agcaggagacccatcatatattatccaagacattacagctcagcagacattgcatg 198  
 DB 62 agcaggagacccatcatatattatccaagacattacagctcagcagacattgcatg 121  
 QY 199 aaagtctaatctctcctcctcctgctgctgacataatgctgacatgctctc 258  
 DB 122 aaagtctaatctctcctcctcctgctgctgacataatgctgacatgctctc 181  
 QY 259 agcagcctgaatccagaggtgcgcagagcagcagcagcagcagcagcagcagc 318  
 DB 182 agcagcctgaatccagaggtgcgcagagcagcagcagcagcagcagcagcagc 241  
 QY 319 tggctccaggaagcgcgcaagaatgagtgcaagaatggtctccgagagcccgaga 378  
 DB 242 tggctccaggaagcgcgcaagaatgagtgcaagaatggtctccgagagcccgaga 301

QY 379 agaaatcatgacagtgctggcgctccaaagaagcagtgccctgtgacatcattcaag 438  
 DB 302 agaaatcatgacagtgctggcgctccaaagaagcagtgccctgtgacatcattcaag 361  
 QY 439 ggcagatggaagaagaacagacacacagcagcagcagcagcagcagcagcagcag 498  
 DB 362 ggcagatggaagaagaacagacacacagcagcagcagcagcagcagcagcagcag 421  
 QY 499 gctctccaggaattctccaacaatgctcagtaagaagcttctgctcctgtgaag 558  
 DB 422 gctctccaggaattctccaacaatgctcagtaagaagcttctgctcctgtgaag 481  
 QY 559 ctctgagcgccactctccaatlaaacattctcagcagcagcagcagcagcagcagcag 618  
 DB 482 ctctgagcgccactctccaatlaaacattctcagcagcagcagcagcagcagcagcag 541  
 QY 619 ccagacactctctctccacactcactccacactcactccacactcactccacactc 678  
 DB 542 ccagacactctctctccacactcactccacactcactccacactcactccacactc 601  
 QY 679 tggctccaaagaatgcttccaagatca 708  
 DB 602 tggctccaaagaatgcttccaagatca 631

RESULT 12

ID AAF68195 standard; cDNA; 533 BP.  
 XX AAF68195;  
 AC AAF68195;  
 XX  
 DT 12-APR-2001 (first entry)  
 XX

DE Human lung tumour protein related nucleotide sequence SEQ ID NO:113.

KW Human; lung cancer; lung tumour; lung tumour protein; gene therapy;  
 KW lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;  
 KW cytoskeletal; antisense inhibition; ss.

OS Homo sapiens.

XX WO200100828-A2.  
 XX 04-JAN-2001.

PF 30-JUN-2000; 2000MO-US18061.  
 XX  
 PR 30-JUN-1999; 99US-0346492.  
 PR 15-OCT-1999; 99US-0419356.  
 PR 17-DEC-1999; 99US-0466867.  
 PR 30-DEC-1999; 99US-0476300.  
 PR 06-MAR-2000; 2000US-0519642.  
 PR 22-MAR-2000; 2000US-0533077.  
 PR 10-APR-2000; 2000US-0546259.  
 PR 27-APR-2000; 2000US-0560406.  
 PR 05-JUN-2000; 2000US-0589184.  
 XX  
 PA (COR-) CORIXA CORP.  
 XX  
 PI Wang T, Bangur GS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;  
 PI Retter MM, Mannion J;  
 PI  
 DR MPI: 2001-071488/08.  
 XX  
 PT Lung tumor-associated proteins and the nucleic acids that encode them,  
 PT useful for preventing, diagnosing and treating lung cancer -  
 XX  
 PS Claim 4: Page 186; 436pp: English.  
 XX  
 CC The present invention describes immunogenic portions of lung tumour-  
 CC associated proteins (I) and the nucleic acids (NAs) that encode them.  
 CC (I) have cytostatic activity and can be used in gene therapy, antisense

inhibition and in vaccines. The NAs and the lung tumour-associated proteins they encode may be used in the prevention, treatment and diagnosis of diseases associated with their inappropriate expression, especially lung cancers. For example, the NAs may be administered to treat diseases by rectifying mutations or deletions in a patient's genome that affect the activity of the protein by expressing inactive proteins or to supplement the patients own production of (1). Additionally, the NAs may be used to produce the lung-tumour associated protein, according to standard recombinant DNA methodology. Conversely, antisense NA molecules may be administered to down regulate protein expression by binding with the cells own genes and preventing their expression. The NA and complementary sequences may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar NA sequences in samples, and hence which patients may be in need of treatment for lung cancer. The (1) may be used as antigens in the production of antibodies and in assays to identify modulators (agonists and antagonists) of the expression and activity of the protein. AAF68093 to AAF68878 and AAF68848 to AAF68678 represent human lung tumour protein related nucleotide and protein sequences which are used in the exemplification of the present invention.

Sequence 533 BP; 152 A; 154 C; 113 G; 114 T; 0 other;

Query Match	58.7%	Score 525.8;	DB 22;	Length 533;
Best Local Similarity	99.6%;	Pred. No. 1.3e-154;		
Matches 527; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

QY	179	ctgacccgaacagttgagatgaaagtcttcaatctctctccctccctcgtgtgtgtgccactaa	238
Db	1	ctgacccgaacagttgagatgaaagtcttcaatctctctccctccctcgtgtgtgtgccactaa	60
QY	239	tgtcgaatgcataatgtctctctacagccttgataccaaagggtgcgcgaagagccaaaggacc	298
Db	61	tgtcgaatgcataatgtctctctacagccttgataccaaagggtgcgcgaagagccaaaggacc	120
QY	299	gaagcgaagctcttagagatgtgtccagaagaagcgcgccaaagaatgtgaatgcaaaagat	358
Db	121	gaagcgaagctcttagagatgtgtccagaagaagcgcgccaaagaatgtgaatgcaaaagat	180
QY	359	ggtctccgaagaccgcgaagaaatcttcaagaagttcttggtgtccaaagaagaagt	418
Db	181	ggtctccgaagaccgcgaagaaatcttcaagaagttcttggtgtccaaagaagaagt	240
QY	419	gcccctgtgcatcttcaagggcaatgttgaagaanaacaaagacccaagagcacacagaa	478
Db	241	gcccctgtgcatcttcaagggcaatgttgaagaanaacaaagacccaagagcacacagaa	300
QY	479	agccaaacaagcattccagaagcctgcgcagaagaattcttcaacaaatgtgcagctagaagct	538
Db	301	agccaaacaagcattccagaagcctgcgcagaagaattcttcaacaaatgtgcagctagaagct	360
QY	539	tgtctgtgcttcttgaagagcctctgaagcgcccaactcttcaathtaaacattccaagcaa	598
Db	361	tgtctgtgcttcttgaagagcctctgaagcgcccaactcttcaathtaaacattccaagcaa	420
QY	599	gaagaacgttagacacacttaccgaacactttcttcccaactcgaatcccaatgttac	658
Db	421	gaagaacgttagacacacttaccgaacactttcttcccaactcgaatcccaatgttac	480
QY	659	ccacccttaaatcatctcagttgcttcaaaagcatgttttccaagatc 707	
Db	481	ccacccttaaatcatctcagttgcttcaaaagcatgttttccaagatc 529	

RESULT	13
AAV69618	
ID	AAV69618 standard; DNA; 511 BP.

AC	AAV69618;
XX	
DT	28-JAN-1999 (first entry)
XX	

DE Human secreted protein gene 8 clone HLHCM89.

KW secreted protein; gene therapy; protein therapy; diagnosis; treatment;  
 KW central nervous system; CNS; immune system; cancer; trauma; liver;  
 KW reproductive disorder; congenital malformation; degenerative disease;  
 KW inflammatory disease; neoplasia; metabolic disorder; testis; placenta;  
 KW brain; T cell; spleen; lung; heart; thymoma; sarcoma; endocrine system;  
 KW endocrinopathy; endocrine polylindular syndrome; endocrinoma; sepsis;  
 KW endocrine ophthalmopathy; osteoclastoma; bacterial infection; bone; ds  
 XX  
 SS Homo sapiens.

	Key	Location/Qualifiers
FH		
FT		
CDS		233..427

```
FT      /product= "secreted protein"  
FT      /transl_except= (pos:344..346, aa:Xaa)  
FT      /transl_except= (pos:383..385, aa:Xaa)
```

```
FT      sig_peptide      233..310
FT      /*tag= b
FT      mat_peptide      311..424
FT      /*tag= c
```

PN W09845712-A2.

PD 15-OCT-1998

07-APR-1998; 98WO-US06801.

30-MAY-1997: 97US-0048184.

PR 08-APR-1997: 97US-0042727

PK	08-APR-1997	9705-0042128
PR	08-APR-1997	9705-0042754

PR	08-APR-1997;	97US-0042825;
PR	30-MAY-1997;	97US-0048068

PR 30-MAY-1997; 9705-0048070.

PA (HUMA-) HUMAN GENOME SCI INC

PI Feng P, Ni J, Rosen CA, H

DR WPI; 1998-594496/50.

XX

PT useful for the diagnosis and treatment of e

PT bacterial infections

Claim 4; Page 107; 142pp; English.

This sequence represents a nucleic acid molecule designated Gene B

CC 209074) which encodes a secreted human protein. This gene is expressed

primarily in lung and to a lesser extent in pancreatic carcinoma and gall bladder and is useful as reagents for differential identification

CC The invention relates

CC which are useful for preventing or ameliorating medical

conditions e.g. by protein or gene therapy. Also pathological conditions can be diagnosed by determining the amount of the new polypeptides in a

CC sample or by determining the presence of mutations in the

cc highly expressed in, and include developing products for the diagnosis of

CC reproductive disorders, cancers, congenital malformations, degenerative

CC diseases in testes, placenta, liver, brain and activated T cells, spleen

CC of the endocrine system or other endocrinopathies, e.g. endocrine

CC polyglutular syndrome, endocrinoma, and endocrine ophthalmopathy,  
CC osteoclastoma and other bone remodelling disorders, bacterial infections  
CC and sepsis. The polypeptides are also useful for identifying their  
CC binding partners.  
XX

Sequence 511 BP; 138 A; 138 C; 125 G; 104 T; 6 other;

Query Match 49.7%; Score 445.2; DB 19; Length 511;  
Best Local Similarity 97.9%; Pred. No. 2,6e-129;  
Matches 466; Conservative 4; Mismatches 4; Indels 2; Gaps 2;

QY 1 ctaatctgttaagtaacagcaagagcgtccactcctgttcttcgcccctaagtga 60  
DB 38 ctaatctgttaagtaacagcaagagcgtccactcctgttcttcgcccctaagtga 97  
QY 61 acgtgtgcttggaactaagcaccagcagctgtatcctcgtactgatacc 120  
DB 98 acgtgtgcttggaactaagcaccagcagctgtatcctcgtactgatacc 157  
QY 121 caggagatcagagcctccagcaggaaccttcattatattctcaagaactacagct 180  
DB 158 caggagatcagagcctccagcaggaaccttcattatattctcaagaactacagct 217  
QY 181 gcaacagacgttgcatgaagtttaattcttccctccctcctggttcgtccactaatg 240  
DB 218 gcaacagacgttgcatgaagtttaattcttccctccctcctggttcgtccactaatg 277  
QY 241 ctgattgtcatatgtctctagcagcctgtatccagaggtgcgcagagccagagacga 300  
DB 278 ctgattgtcatatgtctctagcagcctgtatccagaggtgcgcagagccagagacga 337  
QY 301 ggcacagcttcttaagagatgtctccaggaagcgccagcaaatgtgagcaaatgtg 360  
DB 338 ggcacagcttcttaagagatgtctccaggaagcgccagcaaatgtgagcaaatgtg 397  
QY 361 ttctctgaagcccgagagaagaatttcacagtgcttgcgtgcgcacaagaacagtg 420  
DB 398 ttctctgaagcccgagagaagaatttcacagtgcttgcgtgcgcacaagaacagtg 456  
QY 421 cccgtgtgcatatcaagggcaatgtgagaagaacacacaaagggcacacag 476  
DB 457 cccgtgtgcatatcaagggcaatgtgagaagaacacacaaagggcacacag 511

RESULT 14  
AAx40454  
ID AAx40454 standard; cDNA; 365 BP.  
XX  
AC AAx40454;

XX 18-JUN-1999 (first entry)

XX Human secreted protein 5' EST SEQ ID No: 54.

KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
KW forensic; gene therapy; chromosome mapping; signal peptide; prostate;  
KW upstream regulatory sequence; cytokine activity; cell proliferation;  
KW differentiation; haematopoiesis regulation; tissue growth regulation;  
KW reproductive hormone regulation; chemotactic; chemokine; haemostatic;  
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.

OS Homo sapiens.  
XX  
XX MO9906550-A2.  
XX  
XX 11-FEB-1999.  
XX  
XX 31-JUL-1998; 98WO-IB01232.  
XX  
XX 01-AUG-1997; 97US-0905144.  
XX  
XX (GSEST) GENSET.

XX  
PI Duclert A, Dumas Milne Edwards J, Lacroix B;  
XX WPI: 1999-153780/13.  
DR P-PSDB: AAY11732.  
XX

XX New isolated prostate-derived nucleic acids - used to develop  
PT products which may have cytokine, immune regulatory, haematopoiesis  
PT regulating, anti-inflammatory or tumour inhibition activity  
XX  
PS Claim 1; Page 185; 675bp; English.

XX AAx40438 to AAx40715 represent 5' expressed sequence tags (ESTs) for  
CC human secreted proteins expressed in prostate, and encode the proteins  
CC given in AAY11716 to AAY11993 respectively. The proteins given represent  
CC the signal peptide and an N-terminal fragment of a secreted protein. The  
CC nucleic acid sequences can be used for producing secreted human gene  
CC products. They can also be used to develop products for diagnosis and  
CC therapy. The proteins obtained may have cytokine activity, cell  
CC proliferation and differentiation activity, haematopoiesis regulating  
CC activity, tissue growth regulating activity, reproductive hormone  
CC regulating activity, chemotactic/chemokine activity, haemostatic and  
CC thrombolytic activity, receptor/ligand activity, anti-inflammatory  
CC activity, tumour inhibition activity or other activities. The products  
CC can be used in forensic, gene therapy and chromosome mapping procedures.  
CC The sequences can also be used for obtaining corresponding promoter  
CC sequences. The nucleic acids encoding the signal peptides can be used for  
CC directing extracellular secretion of a polypeptide or the insertion of a  
CC polypeptide into a membrane, or importing a polypeptide into a cell.  
XX

Sequence 365 BP; 91 A; 100 C; 96 G; 78 T; 0 other;

Query Match 40.5%; Score 362.4; DB 20; Length 365;  
Best Local Similarity 99.7%; Pred. No. 2.1e-103;  
Matches 363; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 39 ctgtctgcacctcaaatgtggaacgtggtgcgtgactaagcatagaccacagctga 98  
DB 2 ctgtctgcacctcaaatgtggaacgtggtgcgtgactaagcatagaccacagctga 61  
QY 99 gtaactcgtgactgagtcacccagagatcagagacccctccagcaggaaccttcattat 158  
DB 62 gtaactcgtgactgagtcacccagagatcagagacccctccagcaggaaccttcattat 121  
QY 159 attcttcaagaacttaagcgtgcacgcagcaggttcgtgtaagtcttaattccct 218  
DB 122 attcttcaagaacttaagcgtgcacgcagcaggttcgtgtaagtcttaattccct 181  
QY 219 cctctgtgtgctgacactaatgtctatgtctcctcagcagcgtgaatccagggt 278  
DB 182 cctctgtgtgctgacactaatgtctatgtctcctcagcagcgtgaatccagggt 241  
QY 279 cgcagagagccacagagagcagcaggtctcagagatgtctcagagagcgagcca 338  
DB 242 cgcagagagccacagagagcagcaggtctcagagatgtctcagagagcgagcca 301  
QY 339 agaattgtagtcgcaaatgtgttcctcagagagcccgagagaagaattcagacagtgc 398  
DB 302 agaattgtagtcgcaaatgtgttcctcagagagcccgagagaagaattcagacagtgc 361  
QY 399 tggg 402  
DB 362 tggg 365

RESULT 15  
AA08344  
ID AA08344 standard; DNA; 357 BP.  
XX  
AC AA08344;  
XX  
XX 30-JUN-2000 (first entry)



...

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OK nucleic - nucleic search, using sw model

Run on: September 16, 2002, 00:47:47 ; Search time 3013.46 Seconds  
(without alignments)  
7524.535 Million cell updates/sec

Title: US-09-700-770-2  
Perfect score: 1680  
Sequence: 1 ggtgtgcagataataagtt.....tcaataaacacttcctctgtg 1680

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
EST:\*  
1: em\_estbta:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estcpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: qb\_est1:\*  
10: qb\_est2:\*  
11: qb\_hlc:\*  
12: qb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	835.2	49.7	945	10	BT772722 603053235
2	791.2	47.1	852	10	BT1818205
3	771	45.9	783	10	BT770190
4	762.4	44.5	807	10	BT1823956
5	746.8	44.4	784	10	BT1821790
6	731.8	43.6	826	10	BT529820
7	702.6	41.8	785	10	BT530180
8	684.6	40.8	955	10	BT1821769
9	671.6	40.0	836	10	BT1820886
10	624.2	37.2	660	10	BT548906
11	623.2	37.1	796	10	BT540252
12	619.2	36.9	912	10	BT5484717
13	618	36.8	658	9	BT505896
14	604	36.0	642	10	BT672220
15	598.8	35.6	640	10	BT5482968
16	588.8	35.0	754	10	BT547238
17	587	34.9	602	10	BT530024

18	581.4	34.6	630	10	BT482764
19	559.2	33.3	743	10	BT548174
20	537	32.0	549	10	BT482718
21	521.4	31.0	539	10	BT1824243
22	519.6	30.9	567	10	BT088537
23	519.4	30.9	556	10	BT5933205
24	513.8	30.6	607	10	BT720102
25	511	30.4	525	9	BT602621
26	484.6	28.8	524	10	BT593204
27	473.4	28.2	541	9	BT384404
28	463.2	27.6	469	9	BT1924632
29	443.8	26.4	482	10	BT095122
30	439.8	26.2	465	10	BT154912
31	434.8	25.9	438	9	BT167650
32	430.6	25.6	463	10	BT743981
33	426.4	25.4	1026	10	BT1821994
34	419.4	25.0	421	10	BT1822130
35	417.8	24.9	421	10	BT154920
36	404.8	24.1	433	10	BT826403
37	404	24.0	428	10	BT5932793
38	401.8	23.9	429	9	BT449208
39	392.2	23.3	397	10	BT693858
40	389.4	23.2	576	9	BT363590
41	380.4	22.6	402	10	BT831058
42	373.4	22.2	390	10	BT829540
43	361.4	21.5	363	9	BT167610
44	354.4	21.1	379	10	BT088529
45	353.6	21.0	360	9	BT073612

ALIGNMENTS

RESULT 1  
BT772722  
LOCUS 603053235P1 NIH\_MGC\_122 Homo sapiens cDNA clone IMAGE:5202801 5', mRNA sequence.  
ACCESSION BT772722  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
NIH-MGC http://mgc.nci.nih.gov/  
1 (bases 1 to 945)  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
CONTACT: Robert Strausberg, Ph.D.  
Email: cgapbs-rt@mail.nih.gov  
Tissue Procurement: Life Technologies, Inc.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LLM11508 row: e column: 10  
High quality sequence stop: 856.  
Location/Qualifiers  
1..945  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="5202801"  
/clone\_lib="NIH\_MGC\_122"  
/lab\_host="DH10B"  
/note="Organ: pooled lung and spleen. Vector: pCMV-SPORT6; Site.1: NotI; Site.2: EcoRV (destroyed); RNA source anonymous pool of 24 week female lung, 16 week female spleen, and 20-22 week male spleens. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size





```

|||||
Db 241 CTGCTCAGTGCATGGGGAAAAAGCCAGGAGCATCCCTGTGTGGAGACCTGTGTG 300
QY 354 aacacgcgtctgaagaacatcatctgtgcgaaggtatcatcaagctaacatctccagctg 413
Db 301 AACACCGTCTTAAGACATCATCTGGCGAAGGTATCATACAGCTAACATCTCCACGCTG 360
QY 414 caagtgaaacctcgcccatgtaccagagagctgtatgaatccccctgtgacatggtg 473
Db 361 CAGGTGAACCTCTGGCCATATGACAGAGAGCTGTGTAGTAAAGTCCCTGGACATGTGTG 420
QY 474 gcttgatccaacagccctgtgtcaagaacatcgttgagttccacatgaagcagcagcc 533
Db 421 GCTGATTCACACGCGCCCTGTGTCAAGACATCGTGTGATGCATGCATGACGATGAGGCC 480
QY 534 caagccacatctcgatgtgaacacccagtgaaagtgtgccccacccgctgtgtctcaatgac 593
Db 481 CAAAGCCACCATCCGATGGAGACACCAAGTGCACAGTGGCCACCCGCTGTGTCTCAATGAC 540
QY 594 tgtgcacacagccatgtgagcctgtgcacacacatgtgtcatgaagctctctcctgtgtg 653
Db 541 TGTGGCACCAGCATGSGAGGCTGGGCAATCTGTGTCAATGAGTCTCTCTCCGTGTG 600
QY 654 aacgccttagctaaagcaggtcatgaacctcctagtgccatccctgccaatctagtga 713
Db 601 AAGCGCTTAGCTAAGAGGAGTCAAGACCTCTAGTGCATCCCTGCCAATCATAGTGA 660
QY 714 aaccagctgtgtccggtgtgagaggtcttcaatgagatgtatgaacctccgca- 772
Db 661 AACCGACTGTGTCCCTGTATGAGGCTTCTTCAATGAGATGTCAACACCTCCGCGAG 720
QY 773 gctgtgaaggtgtccatttccctcagcatgtacgcgtcgg-agtgtgacctctgtac 831
Db 721 GCTGGTGAAGGTGGCCATTTCCCTCAGCATGTGACCTCTGGAAGTGTGACTTCTGTATC 780
QY 832 ctgcacatgaaggtgaacacatcagctcactcgtg-gggccaagttgttgtaactcacag 890
Db 781 CTGCGCATCAAGGGGTGACACCATTCAGCTTACTCGTGAAGGGCAAGTGTGGAGCTACCG 840
QY 891 gga 894
Db 841 gga 844

RESULT 3
LOCUS B1770190 783 bp mRNA linear EST 25-SEP-2001
DEFINITION 603053221F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5202893 5',
mRNA sequence.
ACCESSION B1770190
VERSION B1770190.1 GI:15761768
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
NIH-MGC http://mhc.nci.nih.gov/
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@pds-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14M1508 row: 1 column: 06
High quality sequence stop: 782.
FEATURES
source 1..783

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5202893"
/clone_1ib="NIH_MGC_122"
/lab_host="DH10B"
/note="organ: pooled lung and spleen; Vector: pCMV-Sport6;
site_1: NotI; Site-2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung; 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 026. Note:
this is a NIH_MGC Library."
BASE COUNT 167 a 261 c 199 g 156 t
ORIGIN
Query Match 45.9%; Score 771; DB 10; Length 783;
Best Local Similarity 99.9%; Pred. No. 7.4e-176;
Matches 782; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 53 ggaagcggcggagagcctcagcgtgtccaggtctgtgcatctgtgactgtgtccctctga 112
Db 1 GGAGCGGGCGGAGAGACTCAGCGTGGCCAGTGTGAGTCTGACATCTGACTGCTGCTCTGA 60
QY 113 cactctggaagatgtgcgcccgtgtgaacctcaacctctctgtgtgttctgtgcagcca 172
Db 61 CACCTGGGAAGATGGCGGCCCTGGACCTTCACCTCTCTGTGTGTGTGGAGCA 120
QY 173 ccttgatccaagccacccctcagtcacagcagttctcaacctcctgcggcccaaatcatca 232
Db 121 CCTTGATCCAGGACACCTTCAGTCCACATGCACTTCTCATCTCTCGGCCCAAAAGTATGA 180
QY 233 aagaaagcttgacacagagagctgaagagcacacaagccacacagcatctgtcagcagctgc 292
Db 181 AAGAAAGCTGACACAGAGAGCTGAAGAGCACACACACACACAGCATCTGACAGACTGC 240
QY 293 cgtctcagtgatgcattgtcggggaaagccagccggagag-attcctgtgtcggcgagctgt 351
Db 241 CGCTGCTCAAGTGCATGCGGGAAMAGCCGAGGAGCATCCCTGTGTGGAGGCTGTG 300
QY 352 tgaacacgctctcgtgaagacatcatctgtgctgaaggtcatcaagcttaacatctccagc 411
Db 301 TGAACACCGTCTTGAGACACATCATCTGTGTGAGGTCTACAGCTTAACATCTCTCAGC 360
QY 412 tgaaggtgaagccctcgcgcaatgaacagagagctgtgtaagaatcccccttgagatgg 471
Db 361 TGCAGGTGAAGCCCTGGCCATGACACAGAGAGCTGTAGCAAGATCCCTGTGACATGG 420
QY 472 tggctgattcaaacagccctgtgtcaagaacctgtgtgaagtttcaatgaagagactgag 531
Db 421 TGGCTGTGATTCAACAGCCCTGTGTCAAGACCATCTGTGAGTCCCATGACACAGAGG 480
QY 532 cccaagccaacatcgcacatggaacacagtgcaagtgtgccccacccgctgtgtcctagtg 591
Db 481 CCCAAGCCATCTCCGATGAGACACCAAGTGAAGTGGCCACCCGCTGTGTCTAGTG 540
QY 592 actgtgccaacagcattggaagcctgtgcatacctaactgtgtcaataagctctcctcctg 651
Db 541 ACTGTGCCACAGCATTGGAGGCTGGCATCACTGCTGATGATGATGCTCTCTCTG 600
QY 652 tgaagccttagtgaagcaggtgtcaactccttagtgcattccctgcgccaatctagtga 711
Db 601 TGAACGCTTAGCTAAGCAGGTATGAACCTCTTAATGTCATCCCTGCCCCAATCTAGTGA 660
QY 712 aaaaacagctgtgtccggtgatcgaagcttccctaagtgcaatgtatgaagacctctgc 771
Db 661 AAAACAGCTGTCTCCGATGAGAGGCTTCTTCAATGACATGTATGACAGACTCTCTGC 720
QY 772 agctgtgaaggtgccaattccctcagcatgtgacagctgtggaattgacctctgtatc 831

```

Db	721	AGCTGCGTGAAGGTGCCCATTTCCCTCAGACATTCACCCTGTGGACTTTGACCTTCTGTATC	780
Oy	832	ctg 834 	
Db	781	CTG 783	
RESULT	4		
LOCUS	B1823956		
DEFINITION	B1823956	807 bp	mRNA
ACCESSION	603039139F1 NIH_MGC_115	Homo sapiens	linear EST 04-OCT-2001
VERSION	B1823956		cdna clone IMAGE:5180336 5'
KEYWORDS	B1823956.1 GI:15935506		
SOURCE	EST.		
ORGANISM	human.		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominidi; Homo.		
JOURNAL	1 (bases 1 to 807)		
COMMENT	NIH-MGC http://mhc.nci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabbs-remail.nih.gov Tissue Procurement: Life Technologies, Inc. CDNA Library Preparation: Life Technologies, Inc. DNA Sequencing by: The I.M.A.G.E. Consortium (LNLN) Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LLM11449 row: m column: 09 High quality sequence stop: 793. Location/Qualifiers 1..807 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:5180336" /clone_lib="NIH_MGC_115" /lab_host="DH10B" /note="Organ: pooled brain, lung, testis; Vector: pcwv-SPOrtE; site_1: NotI; site_2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH_MGC library."		
BASE COUNT	175 a	269 c	202 g
ORIGIN	161 t		
Query Match	45.4%	Score 762.4;	DB 10; Length 807;
Best Local Similarity	99.5%;	Pred. No. 9.1e-174;	
Matches 796; Conservative	0;	Mismatches 1;	Indels 3; Gaps 3;
OY	59	ggccgaagaccacgaagcgatgcacgaagtcttgacattcactgttgccttcgtcacactg	118
Db	4	GGGGGAGAGACATCCAGCGTGCCAGAGTCTGGCAATCTCAGTCTGCTGCCCTTGACACCTG	63
OY	119	ggaagaatgccgcccgttgacattcaacctctctgtgtttgctggagcacacttga	178
Db	64	GGAAGATGGCCGGCCCGGACGACCTTCACCCCTCTGTGTTGCTTGTGGACCACTTGA	123
OY	179	tccaagcacaccttaatgcccatctgcactgcagttctcatcctcctggccccaaaagtcatcaagaa	238
Db	124	TCCAAGCACACCTTAATGTCACATGTCACATTCTCATTCTCGGCCCAAAAGTCATCAAAGAAA	183
OY	239	agctgacacaggagacttaagacacacaagccacacacatccctgcgagcgtgcgcctgc	298

Db	184	ACGTGACACAGGAGCGTGAAGAGACCAACACGCCACCACAGATCTTCGACGACAGCTGCCCGCTGC	243
QY	299	tcaatgtccatgcggygaaagcccaagccggaag-attccctgtctyggcagcccttgatgaca	357
Db	244	TCAGTGCATGCGGGAAAAAGCCAGCCGGAGGAGCATCCCTGCTGGGCGAGCTGGTGAACA	303
QY	358	ccgtctctaaagacatactctgctctgaaggtcatcaagctcaaatctctcaagctgcaag	417
Db	304	CCGTCTTAAGCAGCATATCTTGCTGGAAGGTCAATCAGCTTAACATCTTCAGCTGAGG	363
QY	418	tgaagccctcggccaattgaccagaagctctgtagtcaaatccctctggagactgtgctg	477
Db	364	TGAAGCCCTCGGCAATGACCAGAGGCTGTATGTAAGATCCCTCTGGACATGTTGGCTG	423
QY	478	gattaaacagccctctggttaagaccatctgtgaggttccaaatgaagactggaagcccaag	537
Db	424	GATTCAACACGCCCTCTGCTGTAAGACCATCTGTGAGTTCCATCATGACGACTGAGGCCCAAG	483
QY	538	ccaccatccgaatgagacacagctgcaagtgtgcccaccgcgctgtgtcttaagtactgtg	597
Db	484	CCACCATTCGGCATGTGACACGACGATGCAAGTGGCCCCACCCGCTGTCTTAAGTACTGTG	543
QY	598	ccaccagccatgggagcctctgcacatccaactgtctgaatgaagctctctctctgtgtagc	657
Db	544	CCACCAGCCATGAGGAGGCTCTGCGCATCCACACTGTGCTGATAGCTCTCTCTGCTGTAAGC	603
QY	658	ccttagttaaagcaagtcatgtaacctctagtctcatccctctgcccacatctatgttataaac	717
Db	604	CCTTAGTAAAGCAGAGGTATGACCTCTCTATGTGCAATCCCTGCCCAATCTATGATAAANAAC	663
QY	718	agctgtctccctgtatctgaagagcttccctcaatgagcatgatatgacagctctctgcaagctg	777
Db	664	AGCTGTCTCCCGTATGAGAGGCTTCTCTCAATGAGCATGTATGTCAGACACTCTCTGACACTG	723
QY	778	tgaaggtgcccatcttccctcagcatctgaacgcttgtaagttgacctcttgatctcgca	837
Db	724	TGAAGGTGCCCATTTCCCTCAGCATTTGACCGCTTGAG-TTGACCTTCTGTATCTTGCA	782
QY	838	tcaaggtgacaccatctag 857	
Db	783	TC-AGGCTGACACCATTTCAAG 801	
RESULT	5		
LOCUS	BI821790	784 bp	mRNA linear EST 04-OCT-2001
DEFINITION	603035865F1 NIH_MGC_115 Homo sapiens CDNA clone IMAGE:5176624 5',		
ACCESSION	BI821790		mRNA sequence.
VERSION	BI821790.1	GI:15933340	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 784)		
JOURNAL	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Stransberg, Ph.D.		
	Email: <a href="mailto:csgrbs-r@mail.nih.gov">csgrbs-r@mail.nih.gov</a>		
	Tissue Procurement: Life Technologies, Inc.		
	cDNA Library Preparation: Life Technologies, Inc.		
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LNLN at:		
	<a href="http://image.llnl.gov">http://image.llnl.gov</a>		
	Plate: L1AM11440 row: b column: 17		
	High quality sequence stop: 763.		
FEATURES	Location/Qualifiers		
source	1..784		
	/organism="Homo sapiens"		





[illegible]

FEATURES	Location/Qualifiers
source	1. .955

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5176856"
/clone_1Id="NIH_MGC_115"
/_clone_host="DH10B"
/note="Organ: pooled brain, lung, testis; Vector:
pCWE-Sport6; Site:1: NotI; Site:2: EcoRV (destroyed); RNA
source anonymous pool of 6 male brains, age range 23-27; 1
male lung, age 27; and 1 male testis, age 69. Library is
oligo-dT primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.8 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
021. Note: this is a NIH_MGC library."

```

Query Match	40.8%	Score 684.6	DB 10	Length 955
Best Local Similarity	93.2%	Pred. No. 6,4e-155		
Matches 749; Conservative	0	Mismatches 50	Indels 5	Gaps 3

OY 115 ccggagaatcctgagccgacccctgaaacctcaacccctctctgtgtttctcggagcaacc 174  
 |||  
 Db 19 CGTGCCTCAGATGACGCGCCCGCCGAGACCTTCAACCTCTCTGTGGTTCTTCGGACCCACC 78  
  
 OY 175 ctgatccaaagccaccctcagtcctccacttcagttctcattccctccggcccaaaagtcacaa 234  
 |||  
 Db 79 TTGATCCAAGCACCCTCAGTCCACTCTCAGTTCATCTCTGGCCAAAATCATCTCAA 138  
  
 OY 235 gaaaagctgacacagagactgaagaccacaacgccaacgaatcctctcagcagactgtcgg 294  
 |||

Db	139	GAAGAAGTGAACACAGAGACTGAAGAGACCACAACGCCACACACTCTCCAGACAGTGGCC	198
QY	295	ctgctcaagtgcacatgcyggaaagccagccggaggt-alcctctgtgcttggcagccctgtgtg	353
Db	199	CTGCTCAAGTGGCATATGCGGGAAAGAGCCAGCGAGGACATCCCTGTTCTGGGAGCGCTGGTG	258
QY	354	aacacgcgtccctgaagagacatcatctcgtgttgaagtgcatacagctcaaatctccagttg	413
Db	259	AACACCGCTCTGTAACACATCATCTGGCTGTGAAGGTCATACAGCTTAACATCTCTCAGGTG	318
QY	414	caagtgaaagccctcggccaatlgacccaagagagctgtcagtacaaatctcccttgaacatgtgtg	473
Db	319	CAGGTGAAGCCCTCGGCCCAATGACCAAGAGAGCGCTGATGCAAGATCCCTCGAGCATGGTGTG	378
QY	474	gctgagatcacaacagccccctgtgtaagaacacatcgttgaagttccacatagaagcatgaaggtcc	533
Db	379	GCTGATTTCAACACGCCCTTGCTGATCAAGACCAATCGTGAGATTTCACATGACGACTGAGGCC	438
QY	534	caagacacacatccgcatlbgacacagtgacaagtgtgccccacccgcctgtgctcagtgac	593
Db	439	CAAGCACACATTCGGCATGTGAGCACCAAGTGCMAAGTGGCCCCACCGCGCTGTCTCAAGTAC	498
QY	594	tgtgcaccacagcccatlbgagacctgtgcacatcaactgtctgacataagctctcctctcgtgtg	653
Db	499	TGTGCCACACACATGAGGAGCGCTGGGCATTCACAACGTGCAATACCTCTCCTCTGGTGTG	558
QY	654	aacgccttaagctaaacagatcatgaacctctcagtgacatccctcggccaatctaaatgtgaa	713
Db	559	AACGCTTAAGCTAAGCAGAGGTATGAACCTCTAGTGGCATCTCCCTGACCAATCTAAGTGA	618
QY	714	aacacagctgtgtcccgatgcagagcttccctcacaatgtgcatgtatgacagacctccgtgag	773
Db	619	AACCAAGCTGTGTCCCGTATCGAGGCTCTCTTCAATGAGCATGATGACAGACTCCTCGAG	678
QY	774	cttgtgtgaaggtgtgccca-cttcctctcaagatttgacccgtctlgagtttgaacctc---tgta	829
Db	679	CTGTGGAAGGAGGCCCATTTTCCCTAGGATTAACCGTGTGGAGTTTGACCTTCTTGATATC	738
QY	830	tctctgcacatcaaggttgacacacatctcaactctaaccttgggggccaagttgttgcagctaca	889
Db	739	CTTGCCATCAAAAGGTGACACCATTCAGAGGCTGTACCAATGGGGGCGCAGTTTGTGTGAGCT	798
QY	890	gggaaaggtgaccaaagtgtgttcaa 913	
Db	799	CCACAGGCGAAAGAGGTGAGACCA 822	

RESULT 9	
BI820886	
LOCUS	836 bp mRNA
DEFINITION	60303924F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5175388 5', mRNA sequence.
ACCESSION	BI820886
VERSION	BI820886.1 GI:15932436
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 836)	NIH-MGC	<a href="http://mgc.ncl.nih.gov/">http://mgc.ncl.nih.gov/</a> .	National Institutes of Health, Mammalian Gene Collection (MGC)	
		Unpublished (1999)		Contact: Robert Strausberg, Ph.D.

Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Life Technologies, Inc.  
cDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
Plate: L16M11436 row: 0 column: 05

FEATURES High quality sequence stop: 750.  
Location/Qualifiers  
source 1..836  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:517538"  
/clone\_1b="NIH\_MGC\_115"  
/lab\_host="DH10B"  
/note="Organ: pooled brain, lung, testis; Vector: PCMV-SPORT6; Site:1: NotI; Site:2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH\_MGC Library."

BASE COUNT 172 a 290 c 210 g 164 t  
ORIGIN

Query Match 40.0%; Score 671.6; DB 10; Length 836;  
Best Local Similarity 95.2%; Pred. No. 8,4e-152;  
Matches 790; Conservative 0; Mismatches 29; Indels 11; Gaps 9;

57 cggcgccgaggaactcagcgtgcccaggtctgcatctgctgctcctgacac 116  
1 cggcgccgaggaactcagcgtgcccaggtctgcatctgctgctcctgacac 60  
117 tgggaagatggccggccgtggaaccttctctgtgtttgtgtgaagcacc 176  
61 tgggaagatggccggccgtggaaccttctctgtgtttgtgtgaagcacc 120  
177 gatcaagcaccctgaatcccaatgacgttctcctgcccacaaatcaataa 236  
121 gatccaaagccacctgacgtccacgtctgctgctgctgctgctgctgct 180  
237 aaagctgaacagagagctggaagagacacaaagcagacatcctgacagct 296  
181 aaagctgaacagagagctggaagagacacaaagcagacatcctgacagct 240  
297 gctcagtgccatgcyggaagaaagccagccgagag-atccctgtgtgtg 355  
241 gctcagtgccatgcyggaagaaagccagccgagag-atccctgtgtgtg 300  
356 cagcgtcctgaagacatcctgctgctgctgctgctgctgctgctgctgct 415  
301 cagcgtcctgaagacatcctgctgctgctgctgctgctgctgctgctgct 360  
416 ggtgaa-gccctggcgaatgacagagctgttag-tcaagatccccctgg--acatgg 471  
361 ggtgaa-gccctggcgaatgacagagctgttag-tcaagatccccctgg--acatgg 420  
472 tggctgattcaaacagccctgtgtaagacatcgtgtgagttccacataagcag 521  
421 tggctgattcaaacagccctgtgtaagacatcgtgtgagttccacataagcag 480  
532 cccaagccacacatccgcatggaacacagttcaagttgcccacccgctggtc 591  
481 cccaagccacacatccgcatggaacacagttcaagttgcccacccgctggtc 540  
592 actg-tgcacagcagatg--gagacctgcaatcgaatgctgctgctgcttcc 648  
541 actg-tgcacagcagatg--gagacctgcaatgctgctgctgcttcc 600  
649 tggtagagccttaagtagcaggtatgaacctctagtgccatccctgccaatctag 708  
601 tggtagagccttaagtagcaggtatgaacctctagtgccatccctgccaatctag 660  
709 tgaataaacagcgtgtgtccgtgacaggtcttccatgaatgcatgtagagccctcc 768  
661 tgaataaacagcgtgtgtccgtgacaggtcttccatgaatgcatgtagagccctcc 720

QY 769 tgcagctgtgtagg-tgcccattccctcagcattgacgctgtgg-agttgacctct 826  
Db 721 tgcagctgtgtagg-tgcccattccctcagcattgacgctgtgg-agttgacctct 780  
QY 827 gtatcctgcatcaagggtgacacacatcagcttactctggtgggagcagt 876  
Db 781 gtatcctgcatccctc-acggtgacacaccttctactctggtgggagcagt 829

RESULT 10  
LOCUS B548906 660 bp mRNA linear EST 04-APR-2001  
DEFINITION 602575739F1 NIH\_MGC\_77 Homo sapiens cDNA clone IMAGE:4703816 5', mRNA sequence.  
ACCESSION B548906  
VERSION B548906.1 GI:13547571  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 660)  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: CLONTECH Laboratories, Inc.  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMU)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMU at: http://image.llnl.gov  
Plate: LCM1542 row: n column: 09  
High quality sequence stop: 660.

#### FEATURES

source 1..660  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4703816"  
/clone\_1b="NIH\_MGC\_77"  
/lab\_host="DH10B (TI phage-resistant)"  
/note="Organ: lung; Vector: pNCR-LIB (Clontech); Site:1: SfiI (ggcgcctggcc); Site:2: SfiI (ggcattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATATATGCG-3' and 3' adaptor sequence: 5'-ATCTAGAGCGCCGAGCGCGCGAGATG-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.9 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC Library."

#### BASE COUNT

144 a 222 c 176 g 118 t

Query Match 37.2%; Score 624.2; DB 10; Length 660;  
Best Local Similarity 98.9%; Pred. No. 2.2e-140;  
Matches 648; Conservative 1; Mismatches 4; Indels 2; Gaps 2;

QY 42 gggagagagtaggagggcgccaggaactcagcgtgcccaggtctgcatcctgca 101  
Db 2 gggagagagtaggagggcgccaggaactcagcgtgcccaggtctgcatcctgca 61  
QY 102 ctgcccctctgacacctgaggagatgycgggcccgtgagacctcaacctctctgtgtt 161  
Db 62 ctgcccctctgacacctgaggagatgycgggcccgtgagacctcaacctctctgtgtt 121  
QY 162 gctggagcagcattgataaagcaccctcagttcccatgacgtctcatctcggccc 221  
Db 122 gctggagcagcattgataaagcaccctcagttcccatgacgtctcatctcggccc 181



42	ggagagagagatgagagggccagagatccaaagtcccaaggtctggactctgcaactg	101
Db	2 ggagagggagagagagggccagagatccaaagtcccaaggtctggactctgcaactg	61
QY	102 cggccctcgaacactctgggaagaatggccggccgtggaaactccaccctctctgtgtt	161
Db	62 ctggccctctgaaacctgggaagaatggccggccgtggaaactccaccctctctgtgtt	121
QY	162 gctggcagcacacttgatccaaagccaccctcagtcaccactgactgattctatcctggccc	221
Db	122 gctggcagcacacttgatccaaagccaccctcagtcaccactgactgattctatcctggccc	181
QY	222 aaaatctcaaaaagaagctgacacagagagctgaagacaaacaaacgcaacagatcct	281
Db	182 AAAAGTATCAAAAGAAAAGCTGACACAGAGCTGAAGACCAACAAGCACCAGATCCT	241
QY	282 gcaagaagctgcccgctgctcagatgccaatgcyggaagaacagccgagag -atccctgtgct	340
Db	242 GCACAGCTGCCGCTGCTGCATCGATCCATCCGGGAAAAGCCACCCGAGCATCCGTGCT	301
QY	341 ggagcagcctgtgaaacacgctccgaaagcacatatactgggttgaaagtcaacaagctaa	400
Db	302 GGGCAGCCTGGTGAACAACCCGCTCGAAGCATATCTGGCTGAAGGTCATCAACAGCTAA	361

Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)  
This sequence was derived from the FA5ESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-QV0-BT0229-251>  
099-038-c016t3-1999-10-256t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 620.





OY 468 atgggtgcgcggattcaacagccgccctggtcgaagaccatcgttgagttccaatgacgaact 527  
 Db 423 atggtgctctggattcaacagcccccctgggtcaagaccatgctggagcttccacatgacgact 482  
 OY 528 gaggcccaagcaaccaclccgcattgacacacattgacaatggtgcccaaccgcctggtccctc 587  
 Db 483 gaggcccaagccaccatccgcattggacacacattgacaaatgg-cccaaccgcctctgggtcttc 541  
 OY 588 atgtactgtgcacacgaagccatgaggagcctgcacatccaaactcgttgataaagctctcttc 647  
 Db 542 attgactgtgcacacagccatggagagccctgcacatccaaactg-tgcataaagctctcttc 600  
 OY 648 ctggtgaacgccttaagctaaagcaggtcattagaacacctcctagtgtg 689  
 Db 601 ctggtgaacgccttaagctaaagcaggtcattagaacacctcctagtgtg 642

RESULT 15  
PCA93069

BG482968

### DEFINITION

ACCESSION

**VERSION**  
**KEYWORDS**

**SOURCE**

NOTES

## REFERENCE

**AUTHORS**  
**TITLE**

**JOURNAL  
COMMENT**

1  
2  
3  
4  
5  
6  
7

## FEATURES

BASE COUNT

ORIGIN

Query Mat  
Best Loca

## Matches

QY 42 5

Db 2 G

•  
•  
•  
•

QY	102	ctgccctctgacaacctggagaagatgtgcgcgcgcgcgtgagaccttaaccctctctgtgttt	165
Db	62	ctgccctctgacacccctgggaatattggccggcccgctggagaccttccaccttctctgtgtgttt	122
QY	162	gtctgcaagccaccttgatccaaagccaccctaaagtcacctgagttctcattcctctgcgc	221
Db	122	gcctggagacgaccttgatctcgaagccaccctcagttccctcagttctcagttctcagttcc	181
QY	222	aaaagtcattcaagaagaagctgtacacagagatctaaagagacacagacacacacacac	281
Db	182	aaaagtcattcaagaagaagaagctgtacacagagatctaaagagacacacacacacacac	241
QY	282	gaagcaagctgtccgcgtctgtcaatgtccatgtcgggaaagccagccggagag -atccctgtgt	340
Db	242	gcagagagctgtccgcgtctgtcaatgtccatgtcgggaaagccagccggagagctgtgtgt	301
QY	341	ggcgaagcctgtgtgaacacacctctctgaagacatatctgtgtgaagtctatcaagctaa	400
Db	302	ggcgaagcctgtgtgaacacacctctctgaagacatatctgtgtgaagtctatcaagctaa	361
QY	401	catcttcagctgtcaagtgtaagccctcgcgcacatgacacagagctgtctgaatcaagctcc	460
Db	362	catcttcagctgtcagctgtgaagccctcgcgcacatgacacagagctgtctgaatcaagctcc	421
QY	461	ccctgacaatgtgtgctgattccaacaagccccctgtgtcaagaacatcgtgtgaattcaaat	520
Db	422	cctgcacactgtgtgctgattccaacaagccccctgtgtcaagaacatcgtgtgaattcaaat	481
QY	521	gaagcactgagcccaagacacacatctgcacatgagacacagctgtcaagtgtcccaaccgcct	580
Db	482	gaagcactgagcccaagacacacatctgcacatgagacacagctgtcaagtgtcccaaccgcct	540
QY	581	ggtctcagtgatcgtgtgcacaaagccatgtggagacctgtgcataccaactgtctgataagt	640
Db	541	ggtctcagtgatcgtgtgcacaaagccatgtggagacctgtgcataccaactgtctgataagt	600
QY	641	ctccttcctgtgtgaac -gccttaagtaagcagtgatcaaa 679	
Db	601	ctccttcctgtgtgaacgccttacttaagcagtgatcaaa 640	

Search completed: September 16, 2002, 01:42:43  
Job time: 3296 sec

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 02:57:19 : Search time 3785.3 Seconds  
(without alignments)  
4947.891 Million cell updates/sec

Title: US-09-700-770-5

Perfect score: 895

Sequence: 1 ctaacgttactgacagc.....aataaattttaatgctc 895

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 segs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_da:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_ov:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_fun:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_mu:\*  
21: em\_or:\*  
22: em\_or:\*  
23: em\_ov:\*  
24: em\_pat:\*  
25: em\_ph:\*  
26: em\_pl:\*  
27: em\_ro:\*  
28: em\_sts:\*  
29: em\_un:\*  
30: em\_un:\*  
31: em\_vl:\*  
32: em\_vl:\*  
33: em\_vl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Length DB	ID	Description
------------	-------------	-----------	----	-------------

1	895	100.0	895	6	AX224644	Sequence
2	895	100.0	1171	6	AX224648	Sequence
3	852	95.2	870	6	AX092294	Sequence
4	525.8	58.7	533	6	AX062486	Sequence
5	475	53.1	626	6	AX182041	Sequence
6	438.4	49.0	168586	9	AC011497	Homo sapi
7	414.8	46.3	419	6	AX332462	Sequence
8	357	39.9	357	6	E40581	Novel prote
9	312	34.9	312	6	AX062416	Sequence
10	291	32.5	291	6	E40584	Novel prote
11	245.6	27.4	357	6	E40583	Novel prote
12	228	25.5	357	6	E40582	Novel prote
13	200.4	22.4	291	6	E40586	Novel prote
14	190	21.2	44332	9	AC005794	Homo sapi
15	189.2	21.1	281	6	E40585	Novel prote
16	138.2	15.4	40520	9	AC011513	Homo sapi
17	118.4	13.2	43593	9	AC022516	Homo sapi
18	78.8	8.8	190860	2	AC073803	Mus muscu
19	78.8	8.8	215234	2	AC073745	Mus muscu
20	73.4	8.2	97	6	AX182077	Sequence
21	56.8	6.3	7218	6	I66494	Sequence 14
22	45.8	5.1	518	6	AX185399	Sequence
23	44.6	5.0	51	6	E40591	Novel prote
24	44	4.9	125020	9	AF429315	Homo sapi
25	44	4.9	209887	2	AF429315	Mus muscu
26	43.6	4.8	234787	2	AC073780	Mus muscu
27	42.8	4.8	456	6	AX187025	Sequence
28	42.4	4.7	209703	2	AL645853	Mus muscu
29	42.4	4.6	163392	2	AC097208	Rattus no
30	41.2	4.6	67356	9	AL161658	Human DNA
31	41.2	4.5	170212	9	AC058822	Homo sapi
32	40.6	4.5	198614	10	AL596127	Mouse DNA
33	40.6	4.5	211844	2	AC048370	Homo sapi
34	40.6	4.4	209156	2	AL607088	Mus muscu
35	39.4	4.4	205150	2	AC005274	Homo sapi
36	39.2	4.4	167159	2	AC073481	Homo sapi
37	39	4.4	196416	9	AC006483	Homo sapi
38	39	4.3	7218	6	I66494	Sequence 14
39	38.6	4.3	152719	2	AC012345	Homo sapi
40	38.6	4.3	155949	9	AC091022	Homo sapi
41	38.6	4.3	156863	2	AC069249	Homo sapi
42	38.6	4.3	161775	9	AC099512	Homo sapi
43	38.6	4.3	187574	9	AC018622	Homo sapi
44	38.6	4.3	213049	2	AC079370	Mus muscu
45	38.6	4.3	213049	2	AC079370	Mus muscu

#### ALIGNMENTS

RESULT 1  
AX224644 LOCUS AX224644 895 bp DNA linear PAT 10-SEP-2001  
DEFINITION Sequence 2 from Patent WO0161055.  
ACCESSION AX224644  
VERSION AX224644.1 GI:15554773

#### KEYWORDS

#### SOURCE

#### ORGANISM

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

Chen, S.Y., Sun, Y. and Macina, R.A.  
Methods for diagnosing, monitoring, staging, imaging and treating  
lung cancer via lung cancer specific genes  
Patent: WO 0161055-A 2 23-AUG-2001;  
diadexus, Inc. (US)

#### FEATURES

#### source

1. 895  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 245 a 250 c 183 g 217 t  
ORIGIN



Db 742 CTCCTCTGATGTCCTTCTCTGTCAGTCCTAGCCCTGCCCTCCCTACCCAGGCT 801  
QY 781 taggcttaactcagaaagatcagaaactgtagctcttagctagtcatttaac 840  
Db 802 TAGGCTTAATTAACCTGAAAGATTCAGGAACGTGAGCTCTAGCTAGTCATTAAAC 861  
QY 841 cttaaatgcatacagaaagtagcaaaacagaagtcataatatctttaagctc 895  
Db 862 CTTAATGCAATCAGAAAGTAGCAACAGCAAGTCATTAATTTTAAATGTC 916

RESULT 3  
AX092294  
LOCUS AX092294 870 bp DNA linear PAT 21-MAR-2001  
DEFINITION Sequence 25 from Patent WO0116318.  
ACCESSION AX092294  
VERSION AX092294.1 GI:13444461  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 870)  
Eaton,D.L., Filvaroff,E., Gerlitsen,M.E., Goddard,A.,  
Godowski,P.J., Grimaldi,C.J., Gurney,A.L., Watanabe,C.K. and  
Wood,W.I.  
TITLE Secreted and transmembrane polypeptides and nucleic acids encoding  
the same  
JOURNAL Patent: WO 0116318-A 25 08-MAR-2001;  
Genentech, Inc. (US)  
FEATURES  
source 1..870  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 251 a 237 c 176 g 206 t  
ORIGIN

Query Match 95.2%; Score 852; DB 6; Length 870;  
Best Local Similarity 100.0%; Pred. No. 2.1e-254;  
Matches 852; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 ctgcgcctcaaatggaacgtgctgagactaaagcagacacagagctgagatc 103  
Db 1 CTCGGCCTCAATGGAAGCGCTGGCTGGACTAAAGCATACACACAGGCTGAGTATC 60  
QY 104 ctgacctgagtcacccagaggaatcagaagcctccagaggaacccctcatatct 163  
Db 61 CTGACCTGAGTCATCCCCAGGGATCAGAGCCTCCAGCAGGAACCTTCATTATTTCT 120  
QY 164 tcaagaacttcaagctgacagacagatgagatgaagttctaattcttccctcc 223  
Db 121 TCAAGCAACTTAAGCTGACCGACGACAGTGGATGAATGTTAATCTTCCCTCTCC 180  
QY 224 tgttgctgcacataatgctgcatgcatgctctcagagcctgaaatccaggggtc 283  
Db 181 TGTTCGTGCATTAATGCTGATGTCATGCTCTCTAGCAGCTGGAATCCAGGGGTCCCA 240  
QY 284 gaggcacagaggaacagagcaggtcttagagatgctcaggaaggcggccaagaat 343  
Db 241 GAGGCCACAGGGAGCGAGCGCAGGCTTCTAGAGATGCTCCAGGAAGCGCGCAAGAT 300  
QY 344 gtgagtgcagaagattggtctcctgagagcccgagagaagaatctcatgacagtgtcggc 403  
Db 301 GTGAGTGCAGAAATGTTGCTCTGAGAGCCCGAGAGAAATTCAGAGACGTCTTGGGC 360  
QY 404 tgcgaagaagcagtgccctgtgcatcttcaagggcaatgtgaagaacaagaagacc 463  
Db 361 TCGCAAGAGAGCAGTCCCTGTGATCATTTCAAGGGCATGTGAAGAAACAGACACC 420  
QY 464 aaagagcacacagaaagccaaacagcaltccagagcctgcagcaatcttcaaacat 523  
Db 421 AAAGGACACCAAGAAAGCCAAACAGCATTTCCAGAGCCTGCCAGCAATTTCTCAACAAAT 480

QY 524 gtcaactaagaagacttgcctgctgccttgytagaagctcagcggccactcttcaat 583  
Db 481 GTCAAGTAAGAAAGCTTTGCTGTGCTTGTGTAGAGCTCTGAGCGCCACTCTTCCAATTA 540  
QY 584 aacattctcagcgaagaagacagtagcacacactccagacactctcttccacactc 643  
Db 541 AACATTCTCAGCCAGAAACAGTAGCGACACTACACACTCTTCTTCCACCTC 600  
QY 644 actctccacagtaccaccacctaaatcaltccagtgctctcaaaaagcattttcaa 703  
Db 601 ACTCTCCCACTGTACCAACCCCTAAATCATTCAGAGTCTCAAAAAGCATGTTTTTCAA 660  
QY 704 gatacttgttgttgcctctcctcagtgctctctctctctctcgtcagcttaacgctg 763  
Db 661 GATCATTTTGTGTGTGCTCTCTAGTGTCTTCTCTGCTGCTGATCTTACCTGTGCTC 720  
QY 764 ctcccttaccagagcttagcttaactaacctgaaagattccaggaagctgagcttct 823  
Db 721 CTCCTTACCCAGAGGCTTAGGCTTAATTACCTGAAGATTCCAGGAACCTGTAGCTTCT 780  
QY 824 agctagtgcatltaaccttaaatgcaatcagaagaagtagcaaacagaagtcataata 883  
Db 781 ACCTAGTGTCAATTAACTTAATGCAATGCAGAAAGTAGCAAAAGAGTCAATTAATA 840  
QY 884 ttttaaatgctc 895  
Db 841 TTTTAAATGTC 852

RESULT 4  
AX062486  
LOCUS AX062486 533 bp DNA linear PAT 24-JAN-2001  
DEFINITION Sequence 113 from Patent WO0100828.  
ACCESSION AX062486  
VERSION AX062486.1 GI:12540361  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 533)  
Wang,T., Bangur,C.S., Lodes,M.J., Fanger,G.R., Vedvick,T.S.,  
Carter,D., Reltter,M.W. and Mannion,J.  
TITLE Compositions and methods for the therapy and diagnosis of lung  
cancer  
JOURNAL Patent: WO 0100828-A 113 04-JAN-2001;  
CORIXA CORPORATION (US)  
FEATURES  
source 1..533  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 152 a 154 c 113 g 114 t  
ORIGIN

Query Match 58.7%; Score 525.8; DB 6; Length 533;  
Best Local Similarity 99.6%; Pred. No. 1.2e-152;  
Matches 527; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 179 ctgcacgacagttgagatgaagttctaattctctccctccctcgttgctgagccactaa 238  
Db 1 CTGCACCGACAGTGGAGATGAAGTTCTTAATCTTCCCTCCTGTTGCTGCCACTAA 60  
QY 239 tctgatlgtcaltgctctcagcagcctgaatccaggggtcgccagagggccaagggacc 298  
Db 61 TCTGTATGTCCATGTGCTCTACAGCCTGAATCCAGAGGCTGCCAGAGGCGCACAGGAGCC 120  
QY 299 gaggcagagcctctaggaatggtctcaggaagggcgccaagaatgtgagtgcagaagat 358  
Db 121 GAGGCCAGGCTTCTTAGAGATGGCTCCAGAAAGCGCGCCAAAGATTTGATGCAAAAGATT 180  
QY 359 gttctctgagagcccgagagaagaatctcatgacagtgtctggtcgtgccaagaagcagt 418

```
Db 181 GGTTCCTGAGAGCCCGAGAGAAATTCATGACAGTGTCTGGCGTGCACAAAGAACAGT 240
QY 419 gcccttgatcatcttcaaggaatgtgaagaacaagaacaaagacacccaaaga 478
Db 241 GCCCTGTGATCATTTCAAGGGCAATGTCAAGAAAACCAAGACCAAGGACCCACAGAA 300
QY 479 agccaaacaagaatccagagcctgcagcaattctcaacaatgtcagcctaagaagc 538
Db 301 AGCCAAACAGACATCCCAAGAGCCTGCCAGCAATTTCTCAACAATGTGAGTAAAGAGCT 360
QY 539 ttgcctgccttgtagagagctcgtgagggccacacttccaattcaaatcagcctaaga 598
Db 361 TTGCTGTGCTTTGTAGAGAGCTCTGAGCCGCCCACTCTTCCAAATTAACATTTCTCAGCCAA 420
QY 599 gaagacagtgagacacactaccagaacactctctccacactcactcctccactctac 658
Db 421 GAAGACAGTGAACACACCTACAGACACACTCTTCTCCACCTACTCTCTCCACTGTAC 480
QY 659 ccacccctaatcatctccagctgtctctcaaaagaatgttttcaagatc 707
Db 481 CCACCCCTTAATCATTCACAGTGTCTCAAAAAGCATGTTTTCAGAGATC 529
```

```
RESULT 5
AX182041
LOCUS AX182041 626 bp DNA linear PAR 06-AUG-2001
DEFINITION Sequence 51 from Patent WO0146697.
ACCESSION AX182041
VERSION AX182041.1 GI:15133314
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 626)
AUTHORS Lille,J., Palermo,A., Wang,Y., Steinmann,K. and Elias,J.
TITLE Identification, assessment, prevention, and therapy of breast
cancer
JOURNAL Patent: WO 0146697-A 51 28-JUN-2001;
Milestone Predictive Medicine, Inc. (US)
FEATURES
source location/Qualifiers
1..626
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 172 a 170 c 149 g 133 t 2 others
ORIGIN
```

```
Query Match 53.1%; Score 475; DB 6; Length 626;
Best Local Similarity 97.1%; Pred. No. 9.1e-137;
Matches 494; Conservative 0; Mismatches 12; Indels 3; Gaps 1;
```

```
QY 57 gggaagcttgctgctgactaaagcatagacacagagctgagctgacctgagctga 116
Db 35 GGGAGACCTGAGCTGGAGCTAAAGCATAGACACAGGCTGATATCTGAGCTGATCA 94
QY 117 tcccaaggatcagagagcctccagcagggaaaccttcattatattttaaagaattac 176
Db 95 TCCCAAGGATTCAGAGAGCTCCAGACGAGACTTCCATTATTTCTTAAGCAACTTAC 154
QY 177 agctgacacagagcttgatgaatgaattctatcttccctcctcctgtgtgctgact 236
Db 155 AGCTGACACGACAGTTCGATGAAAGTTCTTAATCTTCTCCCTCTCTCTGCTGCACT 214
QY 237 aatgcgatgcatatgctctagcagcctgaatccagggtccgacagagccacagga 296
Db 215 AATGCTGATGCTCATGTCTCTTAGCAGCTGATTCAGAGGGGTGCGACAGCCACAGGA 274
QY 297 ccgaagcagagcttcaagagaatgctccaggaagcgagcgaagaatgtgagtgcaaga 356
Db 275 CCGAGGCCAGGCTCTAGAGATGGCTCCAGAGGCGGCCAAGATGTGAGTGCAGAAAG 334
```

```
QY 357 ttgtctctgagagcccgagaagaattcatbaccagtgctggtgctgccaagaagca 416
Db 335 TTGTTCTTGAGAGCCCGAGAGAAATTCATGACAGTGTCTGGGCTGCCAAAGAGCA 394
QY 417 gtccccgtatcatcttcaagggcaatgtgaagaacaagaacaaagagcagcag 476
Db 395 GTGCCCTGTGATCATTTCAAGGGCAATGTGAAGAAACAAAGACCAAGGACCAAG 454
QY 477 aaagccaaacaagaatccagagcctgcagcaattctcaacaatgataagtaagaag 536
Db 455 AAAGCCAAACAGACATTCAGAGAGCTCCAGCAATTTCTCAACAATGTGAGTAAAG 514
QY 537 ctggcctc---tgcttgtagagagcctc 562
Db 515 GCTTTGCTGTGCTTTGTAGAGAACTNT 543
```

```
RESULT 6
AC011497
LOCUS AC011497 168586 bp DNA linear PRI 30-JUN-2000
DEFINITION Homo sapiens chromosome 19 clone CTB-50E14, complete sequence.
ACCESSION AC011497
VERSION AC011497.6 GI:8844110
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 168586)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
2 (bases 1 to 168586)
REFERENCE DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 168586)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (30-JUN-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Jun 30, 2000 this sequence version replaced gi:7711512.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
```

```
FEATURES
source location/Qualifiers
1..168586
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="CTB-50E14"
```

```
BASE COUNT 44482 a 43600 c 40989 g 39515 t
ORIGIN
```

```
Query Match 49.0%; Score 438.4; DB 9; Length 168586;
Best Local Similarity 99.8%; Pred. No. 6.1e-125;
Matches 439; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 456 aagacaccaagacacacagaaagcaacaacatccagagcctgcagcaattct 515
Db 70477 AGGACACCAAGGACACACAGAAAGCAACACATTCACAGAGCTGCCACAAATTTCT 70536
QY 516 caaacaatgtcagtaagaagcttgctcgtcctgtgtagaagctctgagcgcacact 575
Db 70537 CAACCAATGTACGTAAAGACTTGTGCTGCTTGTAGAGCTCTGAGCGCCCACTCT 70596
QY 576 tccaattaaacatctcagcacaagaagacagtgagcacactaccagacactcttctct 635
```



[illegible][illegible]

OY	778	gcttaggcctaattacccgtgaagaaltccaggaaactcgtagcttccttagctagtcatc	837
Dd	119	GCTTAGGCGTTAATTCCTGAAGAATTCCAGAAACTGTAGCTTCCACTAGTCATTC	60
OY	838	aacctaatatcgcaatcaggaaagtgcacaacgaagtcataatatltttaatgctc	895
Dd	59	AACCTTAATATGCMAACAGGAAGATGTAGCAAAACAGAGAGTCATTAATTTTAAATGTC	2
RESULT	8		
E40581	LOCUS	E40581	357 bp DNA linear PAT 31-JAN-2002
DEFINITION	Novel protein and process for producing It.		
ACCESSION	E40581.1 GI:18625110		
VERSION	JP 2000139479-A/1.		
KEYWORDS	Homo sapiens.		
SOURCE	Homo sapiens.		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
REFERENCE	Ito,Y., Ogi,K., Tanaka,H. and Kitada,C. 1 (bases 1 to 357)		
AUTHORS	Novel protein and process for producing it Patent: JP 2000139479-A 1 23-MAY-2000;		
TITLE	TAKEEDA CHEM IND LTD		
JOURNAL	OS Homo sapiens (human) PN JP 2000139479-A/1 PD 23-MAY-2000 PF 02-SEP-1999 JP 1999248414		
COMMENT	PR YASUAKI ITO,KAZUHIRO OGI,HIDEYUKI TANAKA,CHIEKO KITADA PC C12N15/09,A61K31/00,A61K31/00,A61K31/00,A61K31/00,PC A61K31/00, PC A61K31/00,A61K31/00,A61K31/00,A61K31/00,A61K31/00, PC A61K38/00, PC A61K45/00,A61K48/00,C07K14/47,C07K16/18,C12P21/02,G01N33/15, PC G01N33/50// PC A61K39/395,C12P21/08,,C12P21/02,C12R1:91),C12N15/00,A61K37/02 CC FH FT source location/Qualifiers FT		
FEATURES	source location/Qualifiers 1..357 /organism="Homo sapiens" /db_xref="taxon:9606"		
BASE COUNT	102 a 92 c 90 g 73 t		
ORIGIN			
Query Match	39.9% Score 357 DB 6 Length 357		
Best Local Similarity	100.0% Pred. No. 5.6e-100		
Matches 357; Conservative	0 Mismatches 0 Indels 0 Gaps 0		
OY	196	atgaagttcttatcttcttccttccttccttccttccttccttccttccttccttccttc	255
Dd	1	ATGAAGTTCTTATCTCTTCCCTCCTCCTCGTGTCGCCACTATGTCATGTCAMGGTC	60
OY	256	tctagcagcttgtaatcaccagggtgcgccaggccacaaggaccggagccaggctcttag	315
Dd	61	TCTACAGCGCTGAATCCAGGGGTCCGCCAGAGGCCACAGAGACCAGAGGCGCTCTAGG	120
OY	316	agatggtctccagggaaggcgccagaatgtgtgtgtcagaagatlggttctcttggaacccc	375
Dd	121	AGATGCTCCAGGAAGGCGGCCAAGAANTGTGAGTCCAAAGATTTGTTCTCTGAGACCCC	180
OY	376	agaagaaaattatgacagtgctggtgcgcgcaagaagaacagtgccccgtgatcatctc	435
Dd	181	AGAAAGAAAATTATATGACACTGTCTGGGCTGCCCAAAGAACAGCACTGCCCTCTGTATCATTTTC	240
OY	436	aaggccaatgtgaagaaaaaacagaaccaaaagcaccaagaagaagccaaacaagatcc	495
Dd	241	AAGGCAATGTGAAGAAAAACAAGACACCCAAGAGCACACAGAAAGCCAAACAGCATTC	300

**QY**      496 agagcctgccgcaatcttcaaacaatgcagctaagaagcttgctgtcccttg    552  
         |||||  
**Db**      301 AGAGCCTGCCACGAATTTCTCAACAATGTCAGCTAAGAAGCTTGTGCTGCCTTG    357

RESULT 9

LOCUS	312 bp	DNA	PAT 24-JAN-2001
DEFINITION	Sequence 43 from Patent WO0100828.	linear	
ACCESSION	AX062416		
VERSION	AX062416.1		
KEYWORDS	GI:12540291		

Query Match	34.9%	Score 312;	DB 6;	Length 312;
Best Local Similarity	100.0%	Pred. No. 6e-86;		
Matches 312; Conservative	0;	Mismatches	0;	Gaps 0

RESULT 10  
E40584

DEFINITION	291 bp	DNA	linear	PAT 31-JAN-2002
FEATURES				
ORIGIN	1			
ACCESION	E40584			
DESCRIPTION	Novel protein and process for producing it.			

**AUTHORS** Ito, Y., Ogi, K., Tanaka, H. and Kitada, C.  
**TITLE** Novel protein and process for producing it  
**JOURNAL** Patent: JP 2000139479-A 4 23-MAY-2000;  
TAKEDA CHEM IND LTD

PF	02-SEP-1999 JP 1999Z48414
PR	
PI	YASUAKI ITO KAZUHIRO OGI HIDEYUKI TANAKA, CHIEKO KITADA PC
CJ	N15/09, A6IK31/00, A6IK31/00, A6IK31/00, A6IK31/00, A6IK31/00, PC
A6IK31/00,	
PC	A6IK31/00, A6IK31/00, A6IK31/00, A6IK31/00, A6IK31/00, A6IK31/00,
PC	A6IK38/00,
PC	A6IK45/00, A6IK48/00, C07K14/47, C07K16/18, C12P21/02, G01N33/15,
PC	G01N33/50//
CC	A6IK39/39S, C12P21/08, (C12P21/02, C12R1:91), C12N15/00, A6IK37/02
CC	
FH	Key
FTM	Location/Qualifiers
FT	1. .291 /organism='Homo sapiens (human)'.

Query Match	32.5%	Score 291;	DB 6;	Length 291;
Best Local Similarity	100.0%	Pred. NO. 2.1e-79;		
Matches 291;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

## RESULT 11

LOCUS	357 bp	DNA	linear
DEFINITION	Novel protein and process for producing it.		
ACCESSION	E40583		
VERSION	E40583.1	GI:18625112	
KEYWORDS	JP 2000139479-A/3.		
SOURCE	Mus sp.		



```

PC      A61K31/00,A61K31/00,A61K31/00,A61K31/00,A61K31/00,
PC      A61K38/00,
PC      A61K45/00,A61K48/00,C07K14/47,C07K16/18,C12P21/02,G01N33/15,
PC      G01N33/50//
PC      A61K39/395,C12P21/08,(C12P21/02,C12N1.91),C12N15/00,A61K37/02
CC
FH      Key
FH      Location/Qualifiers
FT      source
FT      1..291
FT      /organism="Mus sp. (mouse)".
FEATURES
    source
        1..291
        /organism="Mus sp. (mouse)".
BASE COUNT
    88 a 82 c 80 g 41 t
ORIGIN

Query Match      22.4%; Score 200.4; DB 6; Length 291;
Best Local Similarity 80.7%; Pred. No. 4e-51;
Matches 234; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 262 agcctgaatccaggggtgcgcagagccacagagccagagccagcttctagagatgg 321
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 ACCCCGAACCCAGGGGGTCCCGAAGCCAGCGGGACCAACACCTGCTCTAGAGGTGG 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 322 ctccaggaagcgcccaagaatgtgaaatgtgttctctgagagcccgagaga 381
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 61 CTTCTGGAGGTGGCCAGAAATGTAATGCAAAAGATGTGCTCTCAAGCCCAAGAGGA 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 382 aatcacaagcagtgctcgtgcgcagagccagctccctgtgacatcttaagagcg 441
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 121 AAAGCCACAGCAGTCTGGGGCCACCAAGAAAGAGTCTCCCTGATCAGCTCAAGGGC 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 442 aatgtgaagaagaacagaccacaagagccacagagaagccaaacaagatctcagagcg 501
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 181 AGGGGAAAAAACAAGCACCACCAAAAGCACCACAGAGATGTGCCAAAGACCTCCAGAGCC 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 502 tgcgcgaattctcaaaccaatgtcagtaagaagcttctgctgctt 551
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 241 TGCCAGCAATTTCACAAACGATGTACCTGCGAAGCTTTGGCGCCCTT 290
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
AC005794 44332 bp DNA linear PRI 06-OCT-1998
LOCUS Homo sapiens chromosome 19, cosmid F24866, complete sequence.
DEFINITION AC005794
AC005794 GI:3702272
VERSION AC005794.1
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
    1 (bases 1 to 44332)
    Lamerdin,J.E., McCready,P.M., Skowronski,E., Viswanathan,V.,
    Burkhart-Schultz,K., Gordon,L., Dias,J., Ramirez,M., Stillwagen,S.,
    Phan,H., Velasco,N., Do,L., Regala,W., Terry,A., Gerns,U.,
    Dangnan,L., Erler,A., Christensen,M., Georgescu,A., Avila,J.,
    Liu,S., Attix,C., Andreise,T., Trankheim,M., Amico-Keller,G.,
    Coefield,J., Duarte,S., Lucas,R., Bruce,R., Thomas,P., Quan,G.,
    Kromayashi,B., Arellano,A., Saunders,C., Ow,D., Nolan,M., Trong,S.,
    Kobayashi,A., Olsen,A.S. and Carrero,A.V.
    Sequence analysis of a 2.5 Mb region in 19q13.2 containing a
    clustered CEA/PSG gene family
    Unpublished
    2 (bases 1 to 44332)
    Lamerdin,J.E.
    Direct Submission
    Submitted (06-OCT-1998) Joint Genome Institute, Lawrence Livermore
    National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
    Map and sequence oriented from q centromere to telomere. Cosmid
    F24866 overlaps cosmid R32065 to the left from bases 1 to 5,578 of
    this accession, and overlaps F13825 to the right from bases 35,645

```

```

FEATURES
    source
        to 44,332. Additional chromosome 19 map and sequence information
        are available from: http://www-bio.llnl.gov/genome/genome.html.
        Location/Qualifiers
            1..44332
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="19"
            /map="19q13.2 between BCKDHA and D19S217"
            /clone="F24866"
            /cell_line="GV5HL9-5B"
            /clone_id="LL19NC02 F chromosome 19-specific cosmid
            library"
            /note="Cosmid library constructed at LLNL from flow-sorted
            chromosomes from hybrid 5HL2-B, which carries chromosome
            19 as its only human chromosome."
            599..674
            /rpt_family="MIR"
            complement(1521..1691)
            /note="predicted exon, program: grail2exons_human_1.3,
            frame: 1, quality: excellent, score: 84.000"
            complement(join<1522..1880,2814..2974))
            /product="Human CGM7 gene for nonspecific cross-reacting
            antigen (NCA)"
            complement(<1522..2877)
            /gene="CGM7"
            /note="CGM7"
            /note="cancer/embryonic antigen gene family member 7"
            complement(join<1522..1880,2814..2877))
            /gene="CGM7"
            /note="NCA"
            /codon_start=1
            /product="nonspecific cross-reacting antigen W236
            precursor - human [AA 1-141]"
            /protein_id="AAC62829.1"
            /db_xref="GI:3702273"
            /translation="MGPSAPRGGRHPGGLITLASLTFWHPPTVOFTALPSS
            AAEKDVLLACNISEITQAYWHKGTAEKSPILAGVITDQANIPEAAISGRHTY
            PMSGLFONITLEDAGSYLRITNMSYSDQATGQLHNV"
            complement(1984..2039)
            /rpt_family="(CA)n"
            complement(1984..2039)
            /rpt_family="(CA)n"
            2063..2382
            /rpt_family="LINE2"
            complement(4877..6389)
            /rpt_family="MER65_internal"
            complement(6739..7153)
            /rpt_family="MER65A"
            7570..7673
            /rpt_family="LMC/D"
            complement(7805..8107)
            /rpt_family="ALuDb"
            complement(8288..8583)
            /rpt_family="AluSg"
            complement(8955..9420)
            /rpt_family="MER51B"
            complement(9721..9927)
            /rpt_family="L1PA7"
            complement(9919..12991)
            /rpt_family="L1"
            13161..13329
            /rpt_family="MSTA"
            13331..13604
            /rpt_family="AluSx"
            13613..13831
            /rpt_family="MSTA"
            complement(11573..15821)
            /note="BLASTX similarity to (151..233): match: 0.57,
            score: 4.2e-23; database searched: nr; biliary
            glycoprotein h precursor - human g11179435 (M69176)
            biliary glycoprotein a [Homo sapiens] >gn1PID1d105048
            (D90312)"
            complement(15573..15821)
            /note="BLASTX similarity to (151..233): match: 0.57,
            score: 4.7e-23; database searched: nr; biliary
            glycoprotein i precursor - human g11179437 (M72238)
            misc_feature
            misc_feature
            misc_feature

```

```

misc_feature
complement(15573..15800)
/note="BLASTX similarity to 37204 (248..333): match: 0.32
score: 1.0e-06; database searched: nr; (X16356) TM3-CEA
protein [Homo sapiens]"
misc_feature
complement(15612..15722)
/note="Predicted exon, program: gff12exons_human_1.3,
frame: 2, quality: good, score: 58.000"
repeat_region
/rpt_family="MER49"
16142..16326
/rpt_family="Alu0"
complement(16418..16513)
/note="BLASTX similarity to 437784 (3..34): match: 0.65,
score: 7.2e-08; database searched: nr; (Z21818)
carcinoembryonic antigen [Homo sapiens]"
repeat_region
17280..17362
/rpt_family="FRAM/FAM"
/rpt_family="FRAM/FAM"
17560..17662
/rpt_family="MIR"
17666..17801
/rpt_family="Alu0"
17864..17958
/rpt_family="MIR"
17973..19083
/rpt_family="LIMC4"
19174..19461
/rpt_family="AluY"
complement(19462..19504)
/rpt_family="AT-rich"
19543..19624
/rpt_family="MIR"
19786..20086
/rpt_family="AluX"
complement(20324..20730)
/rpt_family="L1PB2"
complement(20855..21143)
/rpt_family="AluYb"
complement(21173..21261)
/rpt_family="MIR"
21341..21542
/rpt_family="MER20"
complement(22570..22669)
/note="BLASTX similarity to M87707 (105..204): match: 1,
score: 6.2e-33; database searched: nr; Human simple repeat
polymorphism"
repeat_region
complement(22675..22723)
/rpt_family="(CA)n"
22976..23187
/note="Predicted exon, program: gff12exons_human_1.3,
frame: 1, quality: good, score: 62.000"
complement(23278..23613)
/rpt_family="LTR8"
complement(23863..24002)
/rpt_family="LTR8"
complement(24872..25091)
/rpt_family="MER4_internal"
complement(25088..25368)
/rpt_family="MER5_internal"
complement(25389..25768)
/rpt_family="MER5_internal"
26147..26317
/note="BLASTX similarity to Q13946 (7..63): match: 0.78,
score: 5.0e-30; database searched: nr; HIGH-APRINTY
CAMP-SPECIFIC 3',5'-CYCLIC PHOSPHODIESTERASE (HCPI)"
26320..26427
/note="BLASTX similarity to Q13946 (65..100): match: 0.63
score: 5.0e-30; database searched: nr; HIGH-APRINTY
CAMP-SPECIFIC 3',5'-CYCLIC PHOSPHODIESTERASE (HCPI)"
26474..26668
/rpt_family="L1M1/2"
26666..28684

```

	repeat_region	/rpf_family="L1M1/2"	
		28652..30135	
	repeat_region	/rpf_family="L1"	
		complement(30136..30213)	
	repeat_region	/rpf_family="THE1A"	
		30219..36641	
	repeat_region	/rpf_family="L1PA7"	
		complement(36659..36969)	
	repeat_region	/rpf_family="THE1A"	
		complement(36970..37380)	
	repeat_region	/rpf_family="THE1A-internal"	
		complement(37381..37656)	
	repeat_region	/rpf_family="THE1B"	
		37657..39226	
	repeat_region	/rpf_family="L1PB1"	
		complement(39711..40104)	
	repeat_region	/rpf_family="L1PB8"	
		complement(40124..40360)	
	repeat_region	/rpf_family="L1PB5"	
		complement(40626..40743)	
	repeat_region	/rpf_family="L1PB8"	
		40791..40952	
	misc_feature	/rpf_family="M1R"	
		complement(41469..42015)	
	/note="BLASTn similarity to A0108164 (2..548); match:		
Query Match	21.2%; Score 190; DB 9; Length 44332;		
Best Local Similarity	84.9%; Pred. No. 1.6e-47;		
Matches 236; Conservative	0; Mismatches 40; Indels 2; Gaps 2		
OY	1 ctaatcgttaccgtacacagcaa-gaacgcgtcaacctacacgtgttctcgccctcaaatggg	59	
Db	23226 CTAACTGTTACTTAACTAAGTAGGCGCAGCATCACCCTGCTGCTTGCCTCAGATGGA	23167	
OY	60 aacgtgscgtggagactaaagatatgaccacagagctgagtctctgactgagtcattcc	119	
Db	23166 AACACTGACCTGGGACTAAAGTGAAGACACCGGGGTGAGTACCCCTGGGCTGAGGCATCC	23107	
OY	120 ccagagatcagagagctccagcagaggaacctccatatacttctcaagaacttacagc	179	
Db	23106 CCAGAGATCAGGAGGCTCCACACGGGGAACCTTCCA-TACAGTCTCAATGACTTACACG	23048	
OY	180 tgcacgcagattgcgatgaagttcttaattctctccctccctcgtgtgctgcacata	239	
Db	23047 TGCACCAACAATACAAATGAAGTCTTAATCTCTTCCCTCGGTGCTGCCACTAAT	22988	
OY	240 gctgatgtccatggtctctcagcagcctgtaattccaggg	277	
Db	22987 GCTAATGACCATGTCTTACGACGCAAGTCCAGGTG	22950	
RESULT 15			
E40585			
LOCUS	E40585	291 bp	DNA linear PAT 31-JAN-2002
DEFINITION	Novel protein and process for producing it.		
ACCESSION	E40585		
VERSION	E40585.1	GI:18625114	
KEYWORDS	JP 2000139479-A/5.		
SOURCE	Rattus sp.		
ORGANISM	Rattus sp. Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
REFERENCE	1 (bases 1 to 291)		
AUTHORS	Ito,Y., Ogi,K., Tanaka,H. and Kitada,C.		
TITLE	Novel protein and process for producing it		
JOURNAL	Patent: JP 2000139479-A 5 23-MAY-2000; TAKEDA CHEM IND LTD		
COMMENT	OS Rattus sp. (rat) PN JP 2000139479-A/5 PD 23-MAY-2000 PF 02-SEP-1999 JP 1999248414		



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 06:31:07 : Search time 3813.96 Seconds  
(without alignments)  
4910.710 Million cell updates/sec

Title: US-09-700-770-5

Perfect score: 895

Sequence: 1 ctaatcgttctacgaacgc.....ataaatattttaatgctc 895

Scoring table:

OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	895	100.0	895	6	AX224644	Sequence
2	895	100.0	1171	6	AX224648	Sequence
3	852	95.2	870	6	AX092294	Sequence
4	474	53.0	626	6	AX182041	Sequence
5	438	48.9	168586	6	AC011497	Homo sapi
6	427	47.7	533	6	AX062486	Homo sapi
7	357	39.9	357	6	E40581	Novel prote
8	316	35.3	419	6	AX332462	Sequence
9	312	34.9	312	6	AX062416	Sequence
10	291	32.5	291	6	E40584	Novel prote
11	68	7.6	97	6	AX182077	Sequence
12	42	4.7	456	6	AX187025	Sequence
13	42	4.7	518	6	AX185599	Sequence
14	40	4.5	51	6	E40591	Novel prote
15	37	4.1	40520	9	AC011513	Homo sapi
16	36	4.0	43593	9	AC022516	Homo sapi
17	31	3.5	47	6	E40590	Novel prote
18	28	3.1	291	6	E40586	Novel prote
19	28	3.1	357	6	E40583	Novel prote
20	28	3.1	190860	2	AC073803	Mus muscu
21	28	3.1	215234	2	AC073745	Mus muscu
22	26	2.9	40	6	E40588	Novel prote
23	26	2.9	40	6	E40589	Novel prote
24	22	2.5	74433	2	AC015098	Drosophila
25	22	2.5	75470	2	AC036159	Homo sapi
26	22	2.5	152786	2	AP000878	Homo sapi
27	22	2.5	168514	3	AC022350	Drosophila
28	22	2.5	177577	3	AC022349	Drosophila
29	21	2.3	300	14	AF409528	HIV-1 Iso
30	21	2.3	13364	4	AF171666	Bos tauru
31	21	2.3	62871	9	AL356860	Human DNA
32	21	2.3	67340	2	AC100905	Mus muscu
33	21	2.3	101961	9	AL133335	Human DNA
34	21	2.3	113431	2	AC094306	Rattus no
35	21	2.3	125413	2	AC026607	Homo sapi
36	21	2.3	128277	2	AC069107	Homo sapi
37	21	2.3	137127	2	AC069194	Homo sapi
38	21	2.3	149642	2	AP004693	Oryza sat
39	21	2.3	156042	2	AC013614	Homo sapi
40	21	2.3	164396	2	AC024961	Homo sapi
41	21	2.3	164396	2	AC024961	Homo sapi
42	21	2.3	170751	2	AC015742	Homo sapi
43	21	2.3	180868	2	AL672040	Homo sapi
44	21	2.3	183624	30	AC019101	Homo sapi
45	21	2.3	189389	9	CNS05TCE	Human chr

## ALIGNMENTS

RESULT 1  
AX224644 LOCUS AX224644 895 bp DNA linear PAT 10-SEP-2001  
DEFINITION Sequence 2 from Patent WO0161055.  
ACCESSION AX224644  
VERSION AX224644.1 GI:15554773

## SOURCE

ORGANISM human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 895)

## REFERENCE

AUTHORS Chen,S.Y., Sun,Y. and Macina,R.A.  
TITLE Methods for diagnosing, monitoring, staging, imaging and treating  
lung cancer via lung cancer specific genes  
JOURNAL Patent: WO 0161055-A 2 23-AUG-2001;  
diabexus, Inc. (US)

## FEATURES

### source

1. 895  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 245 a 250 c 183 g 217 t  
ORIGIN

Query Match 100.0%; Score 895; DB 6; Length 895;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 895; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctaatctgttacgtaacagcaagacagcgctacacacacgtcttcgcctcctaagatgga 60  
 |||||  
 Db 1 CTAATCTGTAGCTAACAGCAAGACAGCCTACCTGTTCTGCGCCCTCAATGGGA 60  
 QY 61 acgtctgcttgagtaacagcaagcagcagcagcagcagcagcagcagcagcagcagc 120  
 |||||  
 Db 61 ACCTGCTGCTGGAGCTAAAGCATAGACCAAGCCTGATATCTGACCTGATGATCC 120  
 QY 121 caggagatcagagacccacagcagcagcagcagcagcagcagcagcagcagcagc 180  
 |||||  
 Db 121 CAGGGATCAGGAGCCTCCAGGAGGACCTTCATTATTTTAAAGCACTTACAGCT 180  
 QY 181 gacccagacagcttgagaaagcttcaatctctctcctcctcctcctcctcctcctc 240  
 |||||  
 Db 181 GCACCGACAGTTCGATGAAAGTCTAATCTTCCCTCCCTGCTGCTGCTGCTGCTGCT 240  
 QY 241 ctgagtcacatgctctcagcagcagcagcagcagcagcagcagcagcagcagcagc 300  
 |||||  
 Db 241 CTGATGTCATGCTCTTAGCAGCCTGATCCAGGGGTCGCGACAGGACCGCA 300  
 QY 301 ggcacagcttctaggaagcttgctcaggaagcagcagcagcagcagcagcagcagcag 360  
 |||||  
 Db 301 GGCCAGGCTTCTAGAGATGCTCCAGGAAGCGCCCAAGATGATGCAAGATTGG 360  
 QY 361 ttccctgagagcccgagagagaaatctacagctgctcgtcgtcgtcgtcgtcgtcgtc 420  
 |||||  
 Db 361 TTCTGAGAGCCCGAGAGAAATTCATGACAGTCTGGGCTGCCAAAGAGCAGTGC 420  
 QY 421 ccctgtgacatcttcaagagcaatgtaagaaacacacacacacacacacacacacacag 480  
 |||||  
 Db 421 CCTGTGATCATTTCAAGGGCATGTGAAGAAACACACCAAGGACCAAGAGAG 480  
 QY 481 ccaaaacagcattccagagcctgcagcaattctcacaacatgctcagcagcagcagc 540  
 |||||  
 Db 481 CCAAAACAAGCATTCAGAGCCTGCCAGCAATTTCTCAAAAGTCAAGCTTAAGAGCTTT 540  
 QY 541 gctctgcttctgtaggaacctctgagcgcacacacacacacacacacacacacacacag 600  
 |||||  
 Db 541 GCTCTGCTTTGTAAGAGCTTGAAGCCCACTCTTCCAAATTAACAATTCACAGCCAGA 600  
 QY 601 agacagtgagac 660  
 |||||  
 Db 601 AGACAGTGAGCACACCTACAGACACCTCTCTCCACACTCCTCCACACTGTAACC 660  
 QY 661 acccctaataatccacagctctcacaagaacatgcttctcacaacacacacacacacac 720  
 |||||  
 Db 661 ACCCTTAATCATTCAGAGCTGCTCAAAAGCATGTTTTCAGAGATCATTTTGAATGTC 720  
 QY 721 ctctctctagct 780  
 |||||  
 Db 721 CTCTCTCTAGTGCTTCT 780  
 QY 781 taggcttaatactcagaaagcttcagaaacagcagcagcagcagcagcagcagcagcag 840  
 |||||  
 Db 781 TAGGCTTAATTAATCTGAAGATTCAGAGAAAGTGTAGCTTCCAGTGAAGTCAATTAAC 840  
 QY 841 cttaaatgcaatcagaaagcttcagaaacagcagcagcagcagcagcagcagcagcag 895  
 |||||  
 Db 841 CTTAAATGCAATCAGAAAGTACCAACAGAAAGTCAATTAATTAATTAATGTC 895

RESULT 2  
 AX224648  
 LOCUS AX224648 1171 bp DNA linear PAT 10-SEP-2001  
 DEFINITION Sequence 6 from Patent WO0161055.  
 ACCESSION AX224648  
 VERSION AX224648.1 GI:15554775

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## FEATURES

## BASE COUNT

## ORIGIN

## Query Match

## Best Local Similarity

## Matches

## Conservative

## Mismatches

## Indels

## Gaps

## 0;

## human.

## Homo sapiens

## Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## 1 (bases 1 to 1171)

## Chen, S.Y., Sun, Y., and Macina, R.A.

## Methods for diagnosing, monitoring, staging, imaging and treating

## lung cancer via lung cancer specific genes

## Patent: WO 0161055-A 6 23-Aug-2001;

## diadexus, Inc. (US)

## location/Qualifiers

## 1..1171

## /organism="Homo sapiens"

## /db\_xref="taxon:9606"

## BASE COUNT 333 a 320 c 225 g 293 t

## ORIGIN

## Query Match

## Best Local Similarity

## Matches

## Conservative

## Mismatches

## Indels

## Gaps

## 0;

## human.

## Homo sapiens

## Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## 1 (bases 1 to 1171)

## Chen, S.Y., Sun, Y., and Macina, R.A.

## Methods for diagnosing, monitoring, staging, imaging and treating

## lung cancer via lung cancer specific genes

## Patent: WO 0161055-A 6 23-Aug-2001;

## diadexus, Inc. (US)

## location/Qualifiers

## 1..1171

## /organism="Homo sapiens"

## /db\_xref="taxon:9606"

## BASE COUNT 333 a 320 c 225 g 293 t

## ORIGIN

## Query Match

## Best Local Similarity

## Matches

## Conservative

## Mismatches

## Indels

## Gaps

## 0;

## human.

## Homo sapiens

## Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## 1 (bases 1 to 1171)

## Chen, S.Y., Sun, Y., and Macina, R.A.

## Methods for diagnosing, monitoring, staging, imaging and treating

## lung cancer via lung cancer specific genes

## Patent: WO 0161055-A 6 23-Aug-2001;

## diadexus, Inc. (US)

## location/Qualifiers

## 1..1171

## /organism="Homo sapiens"

## /db\_xref="taxon:9606"

## BASE COUNT 333 a 320 c 225 g 293 t

## ORIGIN

## Query Match

## Best Local Similarity

## Matches

## Conservative

## Mismatches

## Indels

## Gaps

## 0;

## human.

## Homo sapiens

## Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## 1 (bases 1 to 1171)

## Chen, S.Y., Sun, Y., and Macina, R.A.

## Methods for diagnosing, monitoring, staging, imaging and treating

## lung cancer via lung cancer specific genes

## Patent: WO 0161055-A 6 23-Aug-2001;

## diadexus, Inc. (US)

## location/Qualifiers

## 1..1171

## /organism="Homo sapiens"

## /db\_xref="taxon:9606"

## BASE COUNT 333 a 320 c 225 g 293 t

## ORIGIN

## Query Match

## Best Local Similarity

## Matches

## Conservative

## Mismatches

## Indels

## Gaps

## 0;

## human.

## Homo sapiens

## Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## 1 (bases 1 to 1171)

## Chen, S.Y., Sun, Y., and Macina, R.A.

## Methods for diagnosing, monitoring, staging, imaging and treating

## lung cancer via lung cancer specific genes

## Patent: WO 0161055-A 6 23-Aug-2001;

## diadexus, Inc. (US)

## location/Qualifiers

## 1..1171

## /organism="Homo sapiens"

## /db\_xref="taxon:9606"

## BASE COUNT 333 a 320 c 225 g 293 t

## ORIGIN

## Query Match

## Best Local Similarity

## Matches

## Conservative

## Mismatches

## Indels

## Gaps

## 0;

## human.

## Homo sapiens

## Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## 1 (bases 1 to 1171)

## Chen, S.Y., Sun, Y., and Macina, R.A.

## Methods for diagnosing, monitoring, staging, imaging and treating

## lung cancer via lung cancer specific genes

## Patent: WO 0161055-A 6 23-Aug-2001;

## diadexus, Inc. (US)

## location/Qualifiers

## 1..1171

## /organism="Homo sapiens"

## /db\_xref="taxon:9606"

## BASE COUNT 333 a 320 c 225 g 293 t

## ORIGIN

## Query Match

## Best Local Similarity

## Matches

## Conservative

## Mismatches

## Indels

## Gaps

## 0;

## human.

## Homo sapiens

## Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## 1 (bases 1 to 1171)

## Chen, S.Y., Sun, Y., and Macina, R.A.

## Methods for diagnosing, monitoring, staging, imaging and treating

## lung cancer via lung cancer specific genes

## Patent: WO 0161055-A 6 23-Aug-2001;

## diadexus, Inc. (US)

## location/Qualifiers

## 1..1171

## /organism="Homo sapiens"

## /db\_xref="taxon:9606"

## BASE COUNT 333 a 320 c 225 g 293 t

## ORIGIN

## Query Match

## Best Local Similarity

## Matches

## Conservative

## Mismatches

## Indels

## Gaps

## 0;

## human.

## Homo sapiens

## Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## 1 (bases 1 to 1171)

## Chen, S.Y., Sun, Y., and Macina, R.A.

## Methods for diagnosing, monitoring, staging, imaging and treating

## lung cancer via lung cancer specific genes

## Patent: WO 0161055-A 6 23-Aug-2001;

## diadexus, Inc. (US)

## location/Qualifiers

## 1..1171

## /organism="Homo sapiens"

## /db\_xref="taxon:9606"

## BASE COUNT 333 a 320 c 225 g 293 t

## ORIGIN

## Query Match

## Best Local Similarity

## Matches

## Conservative

## Mismatches

## Indels

## Gaps

## 0;

## human.

## Homo sapiens

## Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## 1 (bases 1 to 1171)

## Chen, S.Y., Sun, Y., and Macina, R.A.

## Methods for diagnosing, monitoring, staging, imaging and treating

## lung cancer via lung cancer specific genes

## Patent: WO 0161055-A 6 23-Aug-2001;

## diadexus, Inc. (US)

## location/Qualifiers

## 1..1171

## /organism="Homo sapiens"

## /db\_xref="taxon:9606"

## BASE COUNT 333 a 320 c 225 g 293 t

## ORIGIN

## Query Match

## Best Local Similarity

## Matches

## Conservative

## Mismatches

## Indels

## Gaps

## 0;

## human.

## Homo sapiens

## Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## 1 (bases 1 to 1171)

## Chen, S.Y., Sun, Y., and Macina, R.A.

## Methods for diagnosing, monitoring, staging, imaging and treating

## lung cancer via lung cancer specific genes

## Patent: WO 0161055-A 6 23-Aug-2001;

## diadexus, Inc. (US)

## location/Qualifiers

## 1..1171

## /organism="Homo sapiens"

## /db\_xref="taxon:9606"

## BASE COUNT 333 a 320 c 225 g 293 t

## ORIGIN

## Query Match

## Best Local Similarity

## Matches

## Conservative

## Mismatches

## Indels

## Gaps

## 0;

## human.

## Homo sapiens

## Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## 1 (bases 1 to 1171)



Db	742	CTCTCTAGAGTCTCTTCTCTCTGCTACAGTCTTACCGCTGTGCCCTTCCTTACCCAGGCT	801
Qy	781	taggctaattacctaagatccaggaatccaggaactgtagcttcctagctagtgatcattaac	840
Db	802	TAGGCTTAATTACCTGGAAGATTCCAGGAACCTAGCTTCTTACTAGTGTGATTTAAAC	861
Qy	841	cttaaatgcattcaggaagttagcaaacagaagtcataataattttaaatgtc	895
Db	862	CTTAATATCAATCAGGAAGTAGCAAAACAGAGTCAATAAATATTTTAAATGTC	916
RESULT	3		
AX092294	AX092294	870 bp	DNA
LOCUS	Sequence	25 from Patent WO0116318.	linear
DEFINITION	AX092294		
ACCESSION	AX092294.1	GI:13444461	
VERSION			
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. (bases 1 to 870)		
AUTHORS	Eaton,D.L., Filvarov,E., Gerritsen,M.E., Goddard,A., Godowksi,P.J., Grimaldi,C.J., Gurney,A.L., Watanabe,C.K. and Wood,W.I.		
TITLE	Secreted and transmembrane polypeptides and nucleic acids encoding the same		
JOURNAL	Patent: WO 0116318-A 25 08-MAR-2001;		
FEATURES	Genetech, Inc. (US)		
SOURCE	Location/Qualifiers		
	1..870		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
BASE COUNT	251 a 237 c 176 g 206 t		
ORIGIN			
Query Match	95.2%; Score 852; DB 6; Length 870;		
Best Local Similarity	100.0%; Pred. No. 0;		
Matches	852; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Qy	44	ctcgccctcaaatgagaaagctgctgctgagcaatagaccacacagctgagatc	103
Db	1	CTCGCCCTCAATATGGGAAGCGTGGCGTGGAGATAAGCATATACCAACAGCGTGAGTATC	60
Qy	104	ctgacctgaatcatcccccaaggatacgaagcctccagcaaggaaaccttcattatctt	163
Db	61	CTGACCTTAGTATATCCACGAGGATCAGGCGCTCCAGCGGGAACCTTCATTATATCTT	120
Qy	164	tcaagcaactcagctgcaacgacagctgagatgaagtttcaatcttcccccctcc	223
Db	121	TCAACCACTTACACTGCTGACCGACCGAAGTTGGCATGAAAGTTTAATCTTCTCCCTCTCC	180
Qy	224	tgttgcctgccaatatgtcgaatgctcgaatgctcttagcagccttgatccaagggctcgcca	283
Db	181	TGTTGCTGCGACTAATGCTGATGTCCATGCGTCTTAGACGCTGATTCACGAGGTGCGCA	240
Qy	284	gagggccacaaggacccgaagcagcgtcttagagatagtctccagaagggcgccagaat	343
Db	241	GAGGGCCACAGGACCGAGGCCAGGGCTTCTAGAGATGCGCTCCAGGAAGCGGCCAAGAT	300
Qy	344	gtgagtgcaaaagattgtgtctcctgagaagcccgagaagaanaattcatgacagtgtctggcc	403
Db	301	GTGAGTGCAGAGATTGGTTCTCTGAAGACCCCGAGAAAGAAATTCATGACAGTGTCTGGCC	360
Qy	404	tggccaaagaagcagtgccctgtga tcatlttcaagggcaatgtgaagaanaacaagacac	463
Db	361	TGCCAAAGAGACAGTGCCTCTGTGATCATTTTCAAGGCAATGTGAAGAAACAAGACACC	420
Qy	464	aaagccaccacaagaaagccaaacaagcattccagaagctgtccagcaatttctcaaacat	523
Db	421	AAAGCCACCACAGAAAGCCAAACAGCATTCACAGGCTCCGACGCAATTTCTCAACAAT	480

QY	524	gtagcctaagaagcttctgtccttctgttgaagctctagcgcccaactcttccaatla	583
Db	481	gtcacctaaagaagcttttgctgctctgttagagactctgagcgcccaactctccaattga	540
QY	584	aacattctcagcgaagaagacagtgagacacactaccagaactcttcttcccaactc	643
Db	541	aacattctcagcgaagaagacagtgagacacactaccagaactcttcttcccaactc	600
QY	644	actctcccaactctacccaacccctaataatcatctcagctgtctctcaaaaagcatltttca	703
Db	601	actctcccaactctacccaacccctaataatcatctcagctgtctctcaaaaagcatltttca	660
QY	704	gatacttctgttctgtctctctcctcagctgtctctctctctcgcgtcagctcagctgtgac	763
Db	661	gatacttctgttctgttctgtctctcctcagctgtctctctctcgcgtcagctcagctgtgac	720
QY	764	ctccctcaccagaagcttaagcttaatacttaactgaagaatctcagaagaactgtagctctc	823
Db	721	ctccctcaccagaagcttaagcttaatacttaactgaagaatctcagaagaactgtagctctc	780
QY	824	agctagtgcttctaacttaacttaatactcagaagaagtagcaagaagatcaataaata	883
Db	781	agctagtgcttctaacttaacttaatactcagaagaagtagcaagaagaagatcaataaata	840
QY	884	tttccaatgtc 895	
Db	841	tttttaantgtc 852	
RESULT	4		
LOCUS	AX182041	626 bp	DNA
DEFINITION	Sequence 51 from Patent WO0146697.		Linear
ACCESSION	AX182041		
VERSION	AX182041.1	GI:15133314	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 626)		
JOURNAL	Lillie,J., Palermo,A., Wang,Y., Steilmann,K. and Elias,J.		
FEATURES	Identification, assessment, prevention, and therapy of breast cancer		
FEATURES	Patent: WO 0146697-A 51 28-JUN-2001;		
SOURCE	Millennium Predictive Medicine, Inc. (US)		
BASE COUNT	172 a 170 c 149 g 133 t	2 others	
ORIGIN			
Query Match	53.0%; Score 474; DB 6; Length 626;		
Best Local Similarity	100.0%; Pred. No. 1,6e-261;		
Matches 474; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
QY	57	gggaacgctcgtcctgggactaagaatagaccacacagcgtgagatcctcgtacctgagta	116
Db	35	gggaacgctcgtcctgggactaagaatagaccacacagcgtgagatcctcgtacctgagta	94
QY	117	tcaccaagatagagagctccagcagaggaaacttccatatactcttaagcaacttac	176
Db	95	tcaccaagatagagagctccagcagaggaaacttccatatactcttaagcaacttac	154
QY	177	agctcaccagcagctgcagatgaagaatcttaactcttccctccctcctgtcgtccact	236
Db	155	agctcaccagcagctgcagatgaagaatcttaactcttccctccctcctgtcgtccact	214
QY	237	aatgtcatgtccatcgtctctcagcagcctgaatcccaagggtcgcacagagccacaggga	296

Db 215 AATGCTGATGTCATGCTCTCTAGCAGACCTGATCCAGGGGTGCGCAGAGGCCACAGGGA 274  
Qy 297 ccgagccagagctcttagagatggtccaggaagcgccgaagaatgtgagtgaaga 356  
|||||  
Db 275 CGAGGCCAGAGCTTTAGAGATGGCTCCAGGAGGGCCAGAAATGTGAGTGCAGAA 334  
Qy 357 ttggtctctgaagaccccgagaagaatctgacagtgtctggtctgccaagaaga 416  
|||||  
Db 335 TTGGTTCCTGAGAGCCCCGAGAAATTCATGACAGTGTCTGGGCTCCCAAGAGCA 394  
Qy 417 gtgcctctgattcatttaagggaatgtgaagaacaagacacacagccacag 476  
|||||  
Db 395 GTCCCTCTGATCATTTTAAAGGCAATGTGAAGAAACACCAAGGACACACAG 454  
Qy 477 aaagccaaacagcattccagagcctgcagcaatttcacaacatgtcagct 530  
|||||  
Db 455 AAAGCCAAACAGCATTCAGAGCCTGCCAGCAATTTCTCAACATGTGACCT 508

## RESULT 5

LOCUS AC011497 168586 bp DNA linear PRI 30-JUN-2000  
DEFINITION Homo sapiens chromosome 19 clone CTB-50E14, complete sequence.  
AC011497  
VERSION AC011497.6 GI:8844110  
KEYWORDS HTG  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## REFERENCE

1 (bases 1 to 168586)  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## AUTHORS

DOE Joint Genome Institute and Stanford Human Genome Center.

## TITLE

Direct Submission

## JOURNAL

Unpublished

## REFERENCE

2 (bases 1 to 168586)  
DOE Joint Genome Institute.

## AUTHORS

Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

## TITLE

Submitted (30-JUN-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

## JOURNAL

On Jun 30, 2000 this sequence version replaced gi:7711512.

## COMMENT

Draft Sequence Produced by DOE Joint Genome Institute

## FEATURES

Location/Qualifiers

## source

1..168586  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="19"  
/clone="CMB-50E14"

## BASE COUNT

44482 a 43600 c 40985 g 39515 t

## ORIGIN

Query Match 48.9%; Score 438; DB 9; Length 168586;  
Best Local Similarity 100.0%; Pred. No. 7.7e-241;  
Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 458 gacaccaaagacacacagaagaacacacagcattccagagcctgcagcaattctca 517  
|||||  
Db 70479 GACACCAAAGGACACACAGAAACCAACAGCATCCAGAGCTCCAGCAATTTCTCA 70538  
Qy 518 aacaatgcaagaagaagcttgcctgcctctgtgaagctcgaagcgccactcttc 577  
|||||  
Db 70539 AACAAATGTCAGCTTAAGAGCTTGTGCTCTCTTTGTAGAGCTCTGAGCGCCACACTCTTC 70598

Qy 578 caattaacattctcagccaagaagacagtgagcacacctaccagacactctctcc 637  
|||||  
Db 70539 CAATTAACATTTCTCAGCCAGAGAGACAGTGAGCACACACTCTTCTTCTCC 70558  
Qy 638 caccctacatcccccacgttaccaccccctaataatcctcagctgtctcaaaagcatgt 697  
|||||  
Db 70659 CACCTCAGCTCTCCACAGTACCCACCCTTAATATTCACATGCTCTCAAAAACATGTT 70718  
Qy 698 ttcaagaatcatttctgtctctcctcagctgtctctctcctcagctagcc 757  
|||||  
Db 70719 TTTCAGATCATTTTGTGTTGTGCTCTCTAGTGTCTCTCTCTGTCAGTTAGCC 70778  
Qy 758 ttgacctcccccctaccagagcttagccttaattacctgaagaattccaggaactgtag 817  
|||||  
Db 70779 TGTGCTCTCCCTTACCAGGCTTAGGCTTAATACCTGAAAGATTCAGAGAACTGTAG 70838  
Qy 818 ctctcagtagtgatcatttaaccttaatgcaatcaggaagaagagaacagaagca 877  
|||||  
Db 70839 CTTCCTAGCTAGTGTCTATTTAATCAATCAAGAAAGTAGCAAAAGAGAGTCAA 70898  
Qy 878 taatatatttaaatgctc 895  
|||||  
Db 70899 TAAATATTTTAAATGTC 70916

## RESULT 6

LOCUS AX062486 533 bp DNA linear PAT 24-JAN-2001  
DEFINITION Sequence 113 from Patent WO0100828.  
AX062486  
VERSION AX062486.1 GI:12540361  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## REFERENCE

1 (bases 1 to 533)  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## AUTHORS

Wang,T., Bangur,C.S., Lodes,M.J., Fanger,G.R., Vedtick,T.S.,  
Carter,D., Retter,M.W. and Mannion,J.

## TITLE

Compositions and methods for the therapy and diagnosis of lung cancer

## JOURNAL

Patent: WO 0100828-A 113 04-JAN-2001;  
CORIXA CORPORATION (US)

## FEATURES

Location/Qualifiers

## source

1..533  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

## BASE COUNT

152 a 154 c 113 g 114 t

## ORIGIN

Query Match 47.7%; Score 427; DB 6; Length 533;  
Best Local Similarity 99.6%; Pred. No. 2.1e-234;  
Matches 527; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 179 ctgcacccagagctgcagatgaagaagtcctcctccctcctgtgtgtgcacaa 238  
|||||  
Db 1 CTGCACCGACAGTGGAGATGAAGAAAGTTCTATCTCTCCCTCTCTGTTGCTGCACACTAA 60  
Qy 239 tgcgatgtcctatgtctctcctcagagctgaatcagaggggtgcagaaagcagagacc 298  
|||||  
Db 61 TCGTATGTCCTATGCTCTCTACACCTGAATCCAGGGGTGCGCCAGAGGCCACAGGACC 120  
Qy 299 gagccagagctctcctgagagatggtcccgagaagggcgccaagaatgtgagtgaagaatt 358  
|||||  
Db 121 GAGGCCAGGCTTTCTAGAGATGGCTCCAGAAAGGCGCCAGAAAGTGAAGCAAGATT 180  
Qy 339 gttcctcgaagaccccgagaagaatcagagagtgctcgtggtcgtccaaagaagcagt 418  
|||||  
Db 181 GTTCTCTGAGAGCCCGAGAGAAATTCATGACAGTGTCTCTGGGCTGCCAAAGAGAGCT 240  
Qy 419 gccctgtgatcatttcaagggcaatgtgaagaacaagacaccaaagggcacacagaa 478  
|||||

Db	241	GCCTCTGTGATCATTTTCAAGGGCAATGTGAAGAAAAACAAGACACCAAGGACACACAGAA	300
Qy	479	agccaaacaaagcattccagagcctgacagcaattctctcaaaacaaigtcaagtaagaagct	538
Db	301	AGCCAAACAAACATCCAGAGACGCTGCCAGCAATTTCTTAAACAAAGTACGATGAAGACT	360
Qy	539	ttgctctgctcttgttagagactctgagcgcgcactcttccaatlaaacattctcagccaa	598
Db	361	TTGCTCTGCTTTGTGAGAGACTGTGAGGGCCGACACTCTTCCAAATTAACATTTCTCAGCCAA	420
Qy	599	gaagaaagtgagacacactaccagacaactctctctccaccctccactctccactgtac	658
Db	421	GAAGCACTGTGAGCACCCTACAGACACTCTTCTTCTCCACCTCCTCCTCCACTGTAC	480
Qy	659	ccaccctccaaatcattccagctgctctcaaaaagcaatgttttccaagctc	707
Db	481	CCACCCCTAAATTCATTCCTCAGTGCCTCTCAAAAAGCATGTTTTCAAGATC	529
RESULT	7		
LOCUS	E40581	357 bp	DNA Linear PAT 31-JAN-2002
DEFINITION	Novel protein and process for producing it.		
ACCESSION	E40581		
VERSION	E40581.1	GI:18625110	
KEYWORDS	JP 2000139479-A/1.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	Ito, Y., Ogi, K., Tanaka, H. and Kitada, C.		
JOURNAL	Novel protein and process for producing it		
COMMENT	Patent: JP 2000139479-A 1 23-MAY-2000;		
	TAKEEDA CHEM IND LTD		
	OS Homo sapiens (human)		
	PN JP 2000139479-A/1		
	PD 23-MAY-2000		
	PF 02-SEP-1999 JP 1999248414		
	PR		
	PI YASUAKI ITO, KAZUHIRO OGI, HIDEYUKI TANAKA, CHIEKO KITADA		
	C12N15/09, A61K31/00, A61K31/00, A61K31/00, A61K31/00, A61K31/00, PC		
	A61K31/00,		
	PC A61K31/00, A61K31/00, A61K31/00, A61K31/00, A61K31/00, A61K31/00,		
	PC A61K38/00,		
	PC A61K45/00, A61K48/00, C07K14/47, C07K16/18, C12P21/02, G01N33/15,		
	PC G01N33/50//		
	PC A61K39/395, C12P21/08, (C12P21/02, C12R1.91), C12N15/00, A61K37/02		
FEATURES			
Source	FT	Location/Qualifiers	
		1..357	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
BASE COUNT	102 a	92 c	90 g 73 t
ORIGIN			
Query Match	39.9%; Score 357; DB 6; Length 357;		
Best Local Similarity	100.0%; Pident. No. 5.2e-194;		
Matches 357; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
Qy	196	atgaagafttcaatctctctctcctctctctctgttctgctgcacctaattgctatgtccatgtgc	255
Db	1	ATGAAGTCTTAATCTCTCCCTCCCTCGCTGGTGCACATTAATCTGATGTCCATGGTTC	60
Qy	256	tctaacacactgaatccatccaggggtgcagagaagccacaggaacccagagccatctctag	315
Db	61	TCTAACACCTGAATCCACAGGGGTGCCACAGAGCCACAGAGACCCGAGGCCATCTTCTAG	120
Qy	316	agatggtcccgaaagcgcgccaagaatgltgagtcacaaagattggttctctgagagccccc	375

Db	121	AGATGGCTCCAGGAAGGCGCCAAAGTGTGAGTCCAAAGATTGGTTCCTGAGACCCCG	180
QY	376	agaagaanaattatbacaagtgtctggtgcctgcgaagaagaacagtgcccttgatcatctc	435
Db	181	AAAGAAATTTATGATGACATGTCTGTGGCTGCCAAAGAACATGTGCCCTGTGATATATTC	240
QY	436	aaggccaatgtlaaagaaacagaacaccaagaagcaccacagagaagaagccaacagatcc	495
Db	241	AAGGGCAATGTGAGAAAGAAACAAGACCCAAAGACACACAGAAAGCCAAACAGCATTC	300
QY	496	agaagctgcagcaattctcaacaatgtaagtaagaagcttgctctgctcttg	552
Db	301	AGAGCTGCGACCAATTTCTCAACAAATGTCAAGTAAGAACGTTGCTGTGCTTTG	357
RESULT	8		
AX332462/c			
LOCUS	AX332462	419 bp	DNA
DEFINITION	Sequence 2971 from Patent WO0194629.		Linear
ACCESSION	AX332462		
VERSION	AX332462.1	GI:18123096	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
TITLE	Young, P.E., Augustus, M., Carter, K.C., Ehner, R., Endress, G.,		
JOURNAL	Horriigan, S., Soppel, D.R. and Weaver, Z.		
FEATURES	Cancer gene determination and therapeutic screening using signature		
source	gene sets		
BASE COUNT	Patent: WO 0194629-A 2971.13-DEC-2001;		
ORIGIN	Avalon Pharmaceuticals (US)		
	Location/Qualifiers		
	1..419		
	/Organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	121 a 63 c 117 g 118 t		
Query Match	35.3%;	Score 316;	DB 6;
Best Local Similarity	99.5%;	Pred. No. 2.3e-170;	
Matches 416;	Conservative 0;	Mismatches 2;	Indels 0;
			Gaps 0;
QY	478	aagccaacaagcatctccagagcctgcgcagcaattcttcaacaatgtcagctaaagaac	537
Db	419	AAGCCAAACAAACATTCACGAGCTGCGCAATTTCTCAACATGTCAGCTAAGAAC	360
QY	538	tttgctctgctcttgtagagctcttgagcgccactctccaattcaaatcatctcaagca	597
Db	359	TTTTGCTCTGCTTTGTGAGAGCTGTGAGGCGCCACATTTCAATTAAACATTTCTGAGCA	300
QY	598	agaagaagctgagcacaccctaccagacaactctctctccacactcaactctcccaatgta	657
Db	299	AGAAAGACATGAGACACACCTACACAGACACTTCTTCTCCACACTACACTCTCCACTGTA	240
QY	658	cccaaccccaaatcatctccagtgctctcaaaaagatgttttcaagatcatcttggttg	717
Db	239	CCACCCCAATTCATTTCCAGTGTCTCAAAAGCATGTTTTTCAAGATCATTTTGTGTTG	180
QY	718	ttgctctctctagtgctctctctctctcgtcagctcttaagctctgcccctccctaccag	777
Db	179	TTTGCTCTCTGATGTGCTCTTCTCTCTGTCGTCAGTCTTAAAGCTGTGCCCTGCCATTACCG	120
QY	778	gcttaagcttaattacclgaagaatccagaagaactgtagcttccctagctagtgatct	837
Db	119	GCTTAGGCTTAATTAACCTGAAGATTCCAGAAACGTAGCTTCCAGTAGTGTGATTT	60
QY	838	aaccttaaatgcaatcagaaagtgagcaaacgaagtaataatatttttaaatgtc	895
Db	59	AACCTTAATGCAATCAGGAAGTATGCAACAGAAAGTCAATTAATATTTTTTAAATGTC	2

```

RESULT 9
AX062416 312 bp DNA linear PAT 24-JAN-2001
LOCUS Sequence 43 from Patent WO0100828.
DEFINITION AX062416
ACCESSION AX062416
VERSION AX062416.1 GI:12540291
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 312)
AUTHORS Wang, T., Bangur, C.S., Lodes, M.J., Fanger, G.R., Vedvick, T.S.,
Carter, D., Retter, M.W. and Mannion, J.
TITILE Compositions and methods for the therapy and diagnosis of lung
cancer.
JOURNAL Patent: WO 0100828-A 43 04-JAN-2001;
CORIXA CORPORATION (US)
FEATURES
source Location/Qualifiers
1..312
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 100 a 82 c 69 g 61 t
ORIGIN

```

```

Query Match 34.9%; Score 312; DB 6; Length 312;
Best Local Similarity 100.0%; Pred. No. 4.7e-168;
Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 325 caggaaagcgcccaagaatgtgaatgtcaaatgtgttcctctgagagcccgcaagaagaa 384
D 1 CAGGAAGGCGGCCCAAGATGTGAGTGCAGAAATGTTCTCTGAGAGCCCGCAGAGAAAA 60
QY 385 tttctgcaagtgctgtggtgtgcaagaagcagtgccctgtgcatcttcaaggagcaat 444
D 61 TTCAATGACAGTGTGTGGCTGCTCAAGAGAGCGCCCTGTGATTTCAAGGGCAAT 120
QY 445 gtgaagaagaacacacacacacacacacacacacacacacacacacacacacacacacac 504
D 121 GTGAAGAAGAACACACACACACACACACACACACACACACACACACACACACACACACAC 180
QY 505 cagcaatttctcaacaatgtcagctaaagaagcttgcctgtgcttctgtgaagctctga 564
D 181 CAGCAATTTCTCAAAACATGTACGTAAAGAGCTTGTCTGCTGTGTAGAGCTCTGA 240
QY 565 ggcgccactcttcaatcaaatctcagcagaagaagcagtgagacacacacacacacacac 624
D 241 GCGCCACTCTTCCAAATTTCAATTTCTACGCCAAGAGACAGTGAACACACTACACAGAC 300
QY 625 actctctctctc 636
D 301 ACTCTCTCTCTC 312

```

```

RESULT 10
E40584 291 bp DNA linear PAT 31-JAN-2002
LOCUS Novel protein and process for producing it.
DEFINITION E40584
ACCESSION E40584
VERSION E40584.1 GI:18625113
KEYWORDS JP 2000139479-A/4.
SOURCE Homo sapiens.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 291)
AUTHORS Ito, Y., Ogi, K., Tanaka, H. and Kitada, C.
TITILE Novel protein and process for producing it
JOURNAL Patent: JP 2000139479-A 4 23-MAY-2000;
TAKEDA CHEM IND LTD

```

```

COMMENT OS Homo sapiens (human)
PN JP 2000139479-A/4
PD 23-MAY-2000
PF 02-SEP-1999 JP 1999248414
PR
PI YASUAKI ITO, KAZUHIRO OGI, HIDEYUKI TANAKA, CHIERO KITADA
PC C12N15/09,A61K31/00,A61K31/00,A61K31/00,A61K31/00,A61K31/00, PC
A61K31/00,
PC A61K31/00,A61K31/00,A61K31/00,A61K31/00,A61K31/00,A61K31/00,
PC A61K38/00,
PC A61K45/00,A61K48/00,C07K14/47,C07K16/18,C12P21/02,G01N33/15,
PC G01N33/50//
CC A61K39/395,C12P21/08,(C12P21/02,C12R1:91),C12N15/00,A61K37/02
FH
FT Key source Location/Qualifiers
1..291
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 90 a 72 c 79 g 50 t
ORIGIN

```

```

Query Match 32.5%; Score 291; DB 6; Length 291;
Best Local Similarity 100.0%; Pred. No. 6.1e-156;
Matches 291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 262 agcctgaatccaggggtcgcagagggccacagagaccagagccagctctcagagatgg 321
D 1 AGCCTGAATCCAGGGGTGCGCCAGAGGCCACAGAGGCCAGAGCCCTCTAGAGATGG 60
QY 322 ctccaggaagcgcccaagaatgtgagtgcaaatgtgttcctgagagcccgagaaga 381
D 61 CTCAGGAAGCGGCCAGAGATGTGAGTGCAGAAATGTTCTCTGAGAGCCCGAGAGAGA 120
QY 382 aaattcatgacagtgtctgtggtgtgcaagaagcagtgccctgtgataatttcaaggcc 441
D 121 AAATTCATGACAGTGTCTGTGGCTGCTCAAGAGAGACATGTCCTGTGATTTCAAGG 180
QY 442 aatgtgaagaagaacacacacacacacacacacacacacacacacacacacacacacac 501
D 181 AATGTGAAGAAGAACACACACACACACACACACACACACACACACACACACACACACAC 240
QY 502 tgcacgaattctctcaacaatgtcagctaaagaagcttgcctgtgctgttg 552
D 241 TGCCAGCAATTTCTCAAAACATGTACGTAAAGAGCTTGTGCTGCTTTG 291

```

```

RESULT 11
AX182077 97 bp DNA linear PAT 06-AUG-2001
LOCUS Sequence 12 from Patent WO0142792.
DEFINITION AX182077
ACCESSION AX182077
VERSION AX182077.1 GI:15133350
KEYWORDS
SOURCE human.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 97)
AUTHORS Schlegel, R., Deeds, J., Berger, A. and Zhao, X.
TITILE Diagnosis of cervical cancer using marker proteins
JOURNAL Patent: WO 0142792-A 12 14-JUN-2001;
Millennium Predictive Medicine, Inc. (US)
FEATURES
source Location/Qualifiers
1..97
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 18 a 16 c 37 g 24 t 2 others
ORIGIN

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostei  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 1 (passes 1 to 40520)  
 DOE Joint Genome Institute and Stanford Human Genome Center  
 Direct Submission  
 Unpublished



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OK nucleic - nucleic search, using sw model

Run on: September 16, 2002, 06:31:41 ; Search time 390.66 Seconds  
(without alignments)  
3933.442 Million cell updates/sec

Title: US-09-700-770-5  
Perfect score: 895  
Sequence: 1 ctaatctgtactacagc.....aataatattttaatgac 895

Scoring table:  
OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_032802.\*

1: /SIDS1/gcgdata/hold-geneseq/geneqn-emb1/NA1980.DAT:\*  
2: /SIDS1/gcgdata/hold-geneseq/geneqn-emb1/NA1981.DAT:\*  
3: /SIDS1/gcgdata/hold-geneseq/geneqn-emb1/NA1982.DAT:\*  
4: /SIDS1/gcgdata/hold-geneseq/geneqn-emb1/NA1983.DAT:\*  
5: /SIDS1/gcgdata/hold-geneseq/geneqn-emb1/NA1984.DAT:\*  
6: /SIDS1/gcgdata/hold-geneseq/geneqn-emb1/NA1985.DAT:\*  
7: /SIDS1/gcgdata/hold-geneseq/geneqn-emb1/NA1986.DAT:\*  
8: /SIDS1/gcgdata/hold-geneseq/geneqn-emb1/NA1987.DAT:\*  
9: /SIDS1/gcgdata/hold-geneseq/geneqn-emb1/NA1988.DAT:\*  
10: /SIDS1/gcgdata/hold-geneseq/geneqn-emb1/NA1989.DAT:\*  
11: /SIDS1/gcgdata/hold-geneseq/geneqn-emb1/NA1990.DAT:\*  
12: /SIDS1/gcgdata/hold-geneseq/geneqn-emb1/NA1991.DAT:\*  
13: /SIDS1/gcgdata/hold-geneseq/geneqn-emb1/NA1992.DAT:\*  
14: /SIDS1/gcgdata/hold-geneseq/geneqn-emb1/NA1993.DAT:\*  
15: /SIDS1/gcgdata/hold-geneseq/geneqn-emb1/NA1994.DAT:\*  
16: /SIDS1/gcgdata/hold-geneseq/geneqn-emb1/NA1995.DAT:\*  
17: /SIDS1/gcgdata/hold-geneseq/geneqn-emb1/NA1996.DAT:\*  
18: /SIDS1/gcgdata/hold-geneseq/geneqn-emb1/NA1997.DAT:\*  
19: /SIDS1/gcgdata/hold-geneseq/geneqn-emb1/NA1998.DAT:\*  
20: /SIDS1/gcgdata/hold-geneseq/geneqn-emb1/NA1999.DAT:\*  
21: /SIDS1/gcgdata/hold-geneseq/geneqn-emb1/NA2000.DAT:\*  
22: /SIDS1/gcgdata/hold-geneseq/geneqn-emb1/NA2001A.DAT:\*  
23: /SIDS1/gcgdata/hold-geneseq/geneqn-emb1/NA2001B.DAT:\*  
24: /SIDS1/gcgdata/hold-geneseq/geneqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	895	100.0	895	21	AAZ29726
2	895	100.0	895	22	AAH77949
3	895	100.0	1171	22	AAH77951
4	858	95.9	870	21	AAZ98202
5	852	95.2	870	21	AAZ65001
6	852	95.2	870	22	AAZ55994
7	852	95.2	870	22	AAZ92070
8	852	95.2	870	22	AAZ41417
9	807	90.2	1148	22	ABA09257

10	630	70.4	654	21	AAC59829	Human secreted pro
11	593	66.3	849	21	AAAO8343	Human TGC-440 secr
12	427	47.7	533	22	AAZ68195	Human lung tumour
13	357	39.9	357	21	AAAO8344	Human TGC-440 secr
14	335	37.4	365	20	AAZ40454	Human secreted pro
15	312	34.9	312	22	AAZ68125	Human lung tumour
16	307	34.3	511	19	AAZ69618	Human secreted pro
17	291	32.5	291	21	AAAO8345	Mature human TGC-4
18	176	19.7	235	20	AAZ40453	Human secreted pro
19	117	13.1	195	16	AAZ725820	Human gene signatu
20	68	7.6	97	22	AAZ07713	Cervical cancer pr
21	42	4.7	456	22	AAH71445	Human cervical can
22	42	4.7	518	22	AAH70020	Human cervical can
23	40	4.5	51	21	AAAO8356	Human TGC-440 secr
24	31	3.5	47	21	AAAO8355	Human TGC-440 secr
25	28	3.1	291	21	AAAO8351	Mature mouse TGC-4
26	28	3.1	764	21	AAAO8350	Mouse TGC-440 secr
27	28	3.1	764	21	AAAO8349	Mouse TGC-440 secr
28	26	2.9	40	21	AAAO8353	Human TGC-440 secr
29	26	2.9	40	21	AAAO8354	Human TGC-440 secr
30	22	2.5	2831	22	AAZ03409	Human reproductive
31	22	2.5	33249	23	ABZ15430	Drosophila melanog
32	20	2.2	20	21	AAC59888	Oligonucleotide pr
33	20	2.2	20	22	AAH77954	PCR primer for CDN
34	20	2.2	41	22	AAZ28691	Human SNP oligonuc
35	20	2.2	467	22	AAZ13198	Human breast cancer
36	20	2.2	469	22	AAZ22057	Human breast cancer
37	20	2.2	481	22	AAZ22068	Human breast cancer
38	20	2.2	641	22	AAZ13187	Human breast cancer
39	20	2.2	3935	23	ABZ16780	Drosophila melanog
40	20	2.2	7852	22	ABZ14491	Human nervous syst
41	20	2.2	10518	23	ABZ19372	Drosophila melanog
42	20	2.2	10616	22	AAZ65565	Human immune/haema
43	20	2.2	10616	22	AAZ65567	Human immune/haema
44	20	2.2	21732	22	ABZ14492	Human nervous syst
45	20	2.2	21732	22	AAZ81131	Human immune/haema

ALIGNMENTS

RESULT 1	
AAZ29726	AAZ29726 standard; DNA; 895 BP.
ID	AAZ29726
AC	AAZ29726;
XX	
DT	27-MAR-2000 (first entry)
XX	
DE	Human lung specific gene-3.
XX	
KW	Lung Specific Gene; LSG; human; diagnostic marker;
KW	prognosticate; Lung cancer; diagnosis; ds.
XX	
OS	Homo sapiens.
XX	
PN	W09960160-A1.
XX	
PD	25-NOV-1999.
XX	
PF	12-MAY-1999; 99WO-US10344.
XX	
PR	21-MAY-1998; 98US-0086212.
XX	
PA	(DIAD-) DIADEXUS LLC.
XX	
PI	Yang F, Macina RA, Sun Y;
XX	
DR	WPI: 2000-116320/10.
XX	
PT	A new method for diagnosing, monitoring and staging lung cancer
XX	
PS	Example 1; Pages 35-36; 40pp; English.

XX The present sequence is a lung specific gene (LSG) from human  
CC clone ID 126263. The LSG has high level of tissue specificity for lungs  
CC and is overexpressed in cancerous tissues. The sequence serves as a  
CC diagnostic marker for detecting, monitoring, staging and prognosticating  
CC lung cancer. The diagnosis involves comparing levels of LSG in samples  
CC obtained from patient and normal control.  
XX

SQ Sequence 895 BP; 245 A; 250 C; 183 G; 217 T; 0 other;

Query Match 100.0%; Score 895; DB 21; Length 895;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 895; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctaatctgtacgttaacgaagacagcagctacacccctgttctgcccctcaatgga 60  
Db 1 ctaatctgtacgttaacgaagacagcagctacacccctgttctgcccctcaatgga 60  
QY 61 acgctgacctggaactaaagcatagaccacagctgagtaacctgacctgagtcaccc 120  
Db 61 acgctgacctggaactaaagcatagaccacagctgagtaacctgacctgagtcaccc 120  
QY 121 caggatcagagcctccacagcaggaacccctcattatattctcacaagcaacttaagct 180  
Db 121 caggatcagagcctccacagcaggaacccctcattatattctcacaagcaacttaagct 180  
QY 181 gcaccgacagctgcatgaaagtcttaactctcctcctcctcctgctgctgacactaatg 240  
Db 181 gcaccgacagctgcatgaaagtcttaactctcctcctcctcctgctgctgacactaatg 240  
QY 241 ctgagtccatgctctcttaagcagctgaatccagggctgcagcagagccacagggacga 300  
Db 241 ctgagtccatgctctcttaagcagctgaatccagggctgcagcagagccacagggacga 300  
QY 301 ggcacagccttctaagagatgctccacaggaaggcccaagaatgtgagctcaagaattgg 360  
Db 301 ggcacagccttctaagagatgctccacaggaaggcccaagaatgtgagctcaagaattgg 360  
QY 361 ttctctgagagcccccagagaagaataattcatgacagtgctggtgcgcaagaagcagctg 420  
Db 361 ttctctgagagcccccagagaagaataattcatgacagtgctggtgcgcaagaagcagctg 420  
QY 421 cccctgtatcatcttcaagagcgaatgtgaagaagaacacagacacagccacacagaag 480  
Db 421 cccctgtatcatcttcaagagcgaatgtgaagaagaacacagacacagccacacagaag 480  
QY 481 ccaacaacagcatccagagcctgcccagcaattctcacaacatgtcagctaaagaagctt 540  
Db 481 ccaacaacagcatccagagcctgcccagcaattctcacaacatgtcagctaaagaagctt 540  
QY 541 gctctgaccttgaagagcctctgagggccacactctcacaatgaattcagctaaagaagctt 600  
Db 541 gctctgaccttgaagagcctctgagggccacactctcacaatgaattcagctaaagaagctt 600  
QY 601 agacagtgagcagacactacacagacactctctctccacactccacactccacactgla 660  
Db 601 agacagtgagcagacactacacagacactctctctccacactccacactccacactgla 660  
QY 661 acccctaataatcattcagctgctcacaagaagcatgttttcaagatcatcttggttgtg 720  
Db 661 acccctaataatcattcagctgctcacaagaagcatgttttcaagatcatcttggttgtg 720  
QY 721 cctctctagtgtct 780  
Db 721 cctctctagtgtct 780  
QY 781 taggcttaattaccgaaagattccaggaagaactgtagcttccctagctgagtcatttaac 840  
Db 781 taggcttaattaccgaaagattccaggaagaactgtagcttccctagctgagtcatttaac 840  
QY 841 cctaaatgcatcaggaagtagcaacagagatcaataatattttaatgtc 895  
Db 841 cctaaatgcatcaggaagtagcaacagagatcaataatattttaatgtc 895

Db 841 cctaaatgcatcaggaagtagcaacagagatcaataatattttaatgtc 895

RESULT 2

AAH77949

ID AAH77949 standard; DNA; 895 BP.

AC AAH77949;

DT 13-NOV-2001 (first entry)

DE Nucleotide sequence of a human Lng104 polypeptide.

KW Human; Lung cancer specific gene; LSG; Lng104; lung cancer; ss.

XX Homo sapiens.

OS Homo sapiens.

FT Key Location/Qualifiers

FT CDS 196..555

FT /tag= "a"

FT /product= "Lng104"

PN WO200161055-A2.

XX 23-AUG-2001.

PF 20-FEB-2001: 2001WO-US05674.

PR 17-FEB-2000: 2000US-0183188.

PA (DIAD-) DIADEXUS INC.

PI Chen S, Sun Y, Macina RA;

XX MPI: 2001-529917/58.

DR P-PSDB: AAG63977.

PT New lung cancer specific gene for the treatment and diagnosis of lung

PT cancer -

PS Claim 1; Page 113; 11pp; English.

XX The present sequence represents a human lung cancer specific gene (LSG), and encodes a polypeptide designated Lng104. LSGs are useful in the treatment and diagnosis of lung cancer. The treatment of lung cancer comprises the administration of a molecule which down regulates the expression of an LSG. An immune response can be mounted against a target cell expressing an LSG. Identification of potential therapeutic agents for use in imaging and treating lung cancer which comprises screening molecules for an ability to bind to or decrease expression of an LSG relative to LSG in the absence of the agent where the ability of a molecule to bind to the LSG or decrease expression of the LSG is indicative of the molecule being useful in imaging and treating lung cancer.

XX SQ Sequence 895 BP; 245 A; 250 C; 183 G; 217 T; 0 other;

Query Match 100.0%; Score 895; DB 22; Length 895;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 895; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctaatctgtacgttaacgaagacagcagctacacccctgttctgcccctcaatgga 60  
Db 1 ctaatctgtacgttaacgaagacagcagctacacccctgttctgcccctcaatgga 60  
QY 61 acgctgacctggaactaaagcatagaccacagctgagtaacctgacctgagtcaccc 120  
Db 61 acgctgacctggaactaaagcatagaccacagctgagtaacctgacctgagtcaccc 120  
QY 121 caggatcagagcctccacagcaggaacccctcattatattctcacaagcaacttaagct 180  
Db 121 caggatcagagcctccacagcaggaacccctcattatattctcacaagcaacttaagct 180



QY 181 gcacgcagctgcgtgaagaagttcctaattcttccctcctctgtgtgctccactaatg 240  
|||||  
Db 181 gcacgcagctgcgtgaagaagttcctaattcttccctcctctgtgtgctccactaatg 240  
QY 241 ctgattcctcagtgctctcagcagctgaatccaggggtgcagaggccacagagccga 300  
|||||  
Db 241 ctgattcctcagtgctctcagcagctgaatccaggggtgcagaggccacagagccga 300  
QY 301 ggcacagcttctagagatggtccaggaaggcgccaaagaatgtagatgcaagaatttg 360  
|||||  
Db 301 ggcacagcttctagagatggtccaggaaggcgccaaagaatgtagatgcaagaatttg 360  
QY 361 ttctcagagcccgagaaataatcatgacagtgctctgggtgcagaaagacagtgctc 420  
|||||  
Db 361 ttctcagagcccgagaaataatcatgacagtgctctgggtgcagaaagacagtgctc 420  
QY 421 ccctgtgatcatctcaagggtcaatgtgaaagaacaagacacacacagagccacagaaag 480  
|||||  
Db 421 ccctgtgatcatctcaagggtcaatgtgaaagaacaagacacacacagagccacagaaag 480  
QY 481 ccaaaagaagcatccagagcctgcagcaatttctcaacaatgtcagctaaagaagcttt 540  
|||||  
Db 481 ccaaaagaagcatccagagcctgcagcaatttctcaacaatgtcagctaaagaagcttt 540  
QY 541 gctcgtccttctagagagctcagagccacactcttcccaatcaattctcagcagaag 600  
|||||  
Db 541 gctcgtccttctagagagctcagagccacactcttcccaatcaattctcagcagaag 600  
QY 601 agacagctgagcacacacacacacacactcttctcctccacactccacactcagcagcc 660  
|||||  
Db 601 agacagctgagcacacacacacacacactcttctcctccacactccacactcagcagcc 660  
QY 661 acccctaataatcattccagtgctcctcaaaaagcagtcttctcaagatcatttgtgttg 720  
|||||  
Db 661 acccctaataatcattccagtgctcctcaaaaagcagtcttctcaagatcatttgtgttg 720  
QY 721 cctcctcagtgctcctcctcctcctcagctcctcagctcctcctcctcctcagcagct 780  
|||||  
Db 721 cctcctcagtgctcctcctcctcctcagctcctcagctcctcctcctcctcagcagct 780  
QY 781 taggcttaattcactgaaagatccaggaactgtagctcctcagctagtgatgcatctaac 840  
|||||  
Db 781 taggcttaattcactgaaagatccaggaactgtagctcctcagctagtgatgcatctaac 840  
QY 841 cttaaatgcaatcaggaagtagcaaacagaaagtcataataatttttaattgtc 895  
|||||  
Db 841 cttaaatgcaatcaggaagtagcaaacagaaagtcataataatttttaattgtc 895

RESULT 3  
AAH7951  
ID AAH7951 standard; DNA; 1171 BP.  
AC AAH7951;  
XX  
XX  
DT 13-NOV-2001 (first entry)  
XX  
XX Nucleotide sequence of a human Lng104 polypeptide.  
DE  
KW Human; lung cancer specific gene; LSG; Lng104; lung cancer; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX Key Location/Qualifiers  
FH CDS 217..576  
FT /\*tag= a  
FT /product= "Lng104"  
XX  
XX PN WO200161055-A2.  
XX  
XX PD 23-AUG-2001.

PF 20-FEB-2001; 2001WO-US05674.  
XX  
XX 17-FEB-2000; 2000US-0183188.  
XX  
XX (DIAD-) DIADEXUS INC.  
XX  
XX Chen S, Sun Y, MacIna RA;  
XX  
XX MPI: 2001-529917/58.  
DR P-PSDB; AAC63977.  
XX  
XX New lung cancer specific gene for the treatment and diagnosis of lung  
XX cancer -  
XX  
XX Claim 1; Page 117; 119pp; English.  
XX  
XX The present sequence represents a human lung cancer specific gene  
XX (LSG), and encodes a polypeptide designated Lng104. LSGs are useful  
XX in the treatment and diagnosis of lung cancer. The treatment of lung  
XX cancer comprises the administration of a molecule which down regulates  
XX the expression of an LSG. An immune response can be mounted against a  
XX target cell expressing an LSG. Identification of potential therapeutic  
XX agents for use in imaging and treating lung cancer which comprises  
XX screening molecules for an ability to bind to or decrease expression  
XX of an LSG relative to LSG in the absence of the agent where the ability  
XX of a molecule to bind to the LSG or decrease expression of the LSG is  
XX indicative of the molecule being useful in imaging and treating lung  
XX cancer.  
SQ Sequence 1171 BP; 333 A; 320 C; 225 G; 293 T; 0 other;

Query Match 100.0%; Score 895; DB 22; Length 1171;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 895; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctatcgtgtacgttaacagacagacagcgtcacctacactcgttctcgcctcaaatggga 60  
|||||  
Db 22 ctatcgtgtacgttaacagacagacagcgtcacctacactcgttctcgcctcaaatggga 81  
QY 61 acgctgctcgtggaactaaagcatagaccacagctgagatccctcagctgagtcaccc 120  
|||||  
Db 82 acgctgctcgtggaactaaagcatagaccacagctgagatccctcagctgagtcaccc 141  
QY 121 caggatcagagacccctccagcaggaaccttcattatattcttcaagaacttaagct 180  
|||||  
Db 142 caggatcagagacccctccagcaggaaccttcattatattcttcaagaacttaagct 201  
QY 181 gcacgcagctgcgtgaagaagttcctaattcttccctcctcctgtgtgcgcaactatg 240  
|||||  
Db 202 gcacgcagctgcgtgaagaagttcctaattcttccctcctcctgtgtgcgcaactatg 261  
QY 241 ctgattcctcagtgctctcagcagctgaatccaggggtgcagaggccacagagccga 300  
|||||  
Db 262 ctgattcctcagtgctctcagcagctgaatccaggggtgcagaggccacagagccga 321  
QY 301 ggcacagcttctagagatggtcccaaggagcgccaaagaatgtagatgcaagaatttg 360  
|||||  
Db 322 ggcacagcttctagagatggtcccaaggagcgccaaagaatgtagatgcaagaatttg 381  
QY 361 ttctcagagcccgagaaataatcatgacagtgctcgtggtgcagaaagacagtgctc 420  
|||||  
Db 382 ttctcagagcccgagaaataatcatgacagtgctcgtggtgcagaaagacagtgctc 441  
QY 421 ccctgtgatcatctcaagggtcaatgtgaaagaacaagacacacacagagccacagaaag 480  
|||||  
Db 442 ccctgtgatcatctcaagggtcaatgtgaaagaacaagacacacacagagccacagaaag 501  
QY 481 ccaaaagaagcatccagagcctgcagcaatttctcaacaatgtcagctaaagaagcttt 540  
|||||  
Db 502 ccaaaagaagcatccagagcctgcagcaatttctcaacaatgtcagctaaagaagcttt 561  
QY 541 gctcgtccttctagagagctcagagccacactcttcccaatcaattctcagcagaag 600

Db 562 gctcgcctttagagagctcagcgccactcttccaattaaattccagcaaga 621  
 Qy 601 aagacgtgaacacactccagacactcttctcccaactactctccacgtacc 660  
 Db 622 agacagtgagcaccctccagacactcttctcccaactactctccacgtacc 681  
 Qy 661 acccctaattcattcagctgcctcacaagaatgttttccaagaattttgtgtg 720  
 Db 682 acccctaattcattcagctgcctcacaagaatgttttccaagaattttgtgtg 741  
 Qy 721 cctctcagtgctctctctctcgtcagctttagcctgtccctccctaccaggct 780  
 Db 742 cctctcagtgctctctctctcgtcagctttagcctgtccctccctaccaggct 801  
 Qy 781 tagcttaattacctgaagaattcagaagaacttagcttctcgtcgtatgtcattaac 840  
 Db 802 tagcttaattacctgaagaattcagaagaacttagcttctcgtcgtatgtcattaac 861  
 Qy 841 cttaaatgcaatcaggaagtagcaacaagaagtcataatattttaatgtc 895  
 Db 862 cttaaatgcaatcaggaagtagcaacaagaagtcataatattttaatgtc 916

## RESULT 4

AAZ98202

ID AAZ98202 standard; cDNA; 870 BP.

AC AAZ98202;

DT 11-MAY-2000 (first entry)

DE Human signal peptide containing protein HSP90-94 cDNA SEQ ID NO:228.

KW Human: signal peptide-containing protein; HSP90; diagnosis; cancer;  
 KW Inflammation; cardiovascular disease; anticancer; anti-inflammation;  
 KW antimicrobial; neurotrophic; neuroprotective; cardiovascular; hepatotropic;  
 KW antiaesthetic; gene therapy; cell proliferation; neurological disorder;  
 KW reproductive disorder; developmental disorder; arteriosclerosis;  
 KW cirrhosis; psoriasis; acquired immune deficiency syndrome; anemia;  
 KW asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;  
 KW Parkinson's disease; Huntington's disease; ovulatory defect;  
 KW muscular dystrophy; ss.

XX Homo sapiens.

XX WO200000610-A2.

XX 06-JAN-2000.

XX 25-JUN-1999; 99MO-US14484.

XX 26-JUN-1998; 98US-0090762.

XX 31-JUL-1998; 98US-0094983.

XX 01-OCT-1998; 98US-0102686.

XX 11-DEC-1998; 98US-0112129.

XX (INCY-) INCYTE PHARM INC.

XX Lal P, Tang YT, Gorgone GA, Corley NC, Guegler KJ, Baughn MR;

XX PI Akerblom IE, Au-Young J, Yue H, Patterson C, Reddy R, Hillman JL;

XX Bandman O;

XX MPI; 2000-160673/14.

XX P-PSDB; AAY87317.

XX New human signal peptide-containing proteins useful in treatment,

XX PT prevention and diagnosis of e.g. cancer, inflammation and

XX PT cardiovascular disease

XX PS Claim 9; Page 303; 327pp; English.

XX CC AAZ98109 to AAZ98242 encode AAY87224 to AAY87357 which represent the

CC human signal peptide-containing proteins HSP90-1 to HSP90-134. HSPs have  
 CC anticancer, anti-inflammatory, antimicrobial, neurotrophic, hepatotropic,  
 CC neuroprotective, cardiovascular and antiaesthetic activities, and can  
 CC be used in gene therapy. HSPs can be used to treat or prevent disorders  
 CC associated with decreased activity or function of HSP. Antagonists of  
 CC HSP are used to treat or prevent disorders associated with increased  
 CC activity or function of HSP. Such diseases include cell proliferation  
 CC (including cancer), inflammation, cardiovascular, neurological,  
 CC reproductive or developmental disorders, (e.g. arteriosclerosis,  
 CC cirrhosis, psoriasis, acquired immune deficiency syndrome, anemia,  
 CC asthma, Crohn's disease, microbial or other infections, congestive or  
 CC ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's  
 CC diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSP  
 CC nucleic acids can be used for the recombinant production of HSP, for  
 CC detecting HSP in standard hybridisation and amplification assays (for  
 CC diagnosis and monitoring), in gene therapy, as antisense,  
 CC triplex-forming or ribozyme therapeutics, for detecting related sequences  
 CC or genetic variations, and for chromosomal mapping. HSP are also used to  
 CC raise specific antibodies (Ab) and to screen for agonists and  
 CC antagonists (potential therapeutic agents). Ab are used to diagnose, or  
 CC monitor, HSP-related diseases (in usual immunoassays), as therapeutic  
 CC antagonists, in competitive drug screens, and for purification of HSP  
 CC from natural sources.

SQ Sequence 870 BP; 244 A; 240 C; 177 G; 209 T; 0 other;

Query Match 95.9%; Score 858; DB 21; Length 870;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 858; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 cctgttcgcctcctaattggaagcctgctcgggacctaagaacatagaccaggctg 97  
 Db 1 cctgttcgcctcctaattggaagcctgctcgggacctaagaacatagaccaggctg 60  
 Qy 98 agtatcctgacctgagtcaccccaaggatcaggaagcctccaggaagccttcatt 157  
 Db 61 agtatcctgacctgagtcaccccaaggatcaggaagcctccaggaagccttcatt 120  
 Qy 158 tatcttcaagaacttaagcctgcgcagcagcttgcatgtaagttaactctctcc 217  
 Db 121 tatcttcaagaacttaagcctgcgcagcagcttgcatgtaagttaactctctcc 180  
 Qy 218 tctctcgtgtgctcactaatgtatgtcattcattcattcattcattcattcattc 277  
 Db 181 tctctcgtgtgctcactaatgtatgtcattcattcattcattcattcattcattc 240  
 Qy 278 tcgcagagggccacagggcagcagcagcagcagcagcagcagcagcagcagcagc 337  
 Db 241 tcgcagagggccacagggcagcagcagcagcagcagcagcagcagcagcagcagc 300  
 Qy 338 aagaatgtagtgcaagaattgtctcctttagagcccggaagaagaatcagcagctg 397  
 Db 301 aagaatgtagtgcaagaattgtctcctttagagcccggaagaagaatcagcagctg 360  
 Qy 398 ctgggctgcacaaagaagcagctgcccctgtgacattcacaaggcaatgtgaagaacaa 457  
 Db 361 ctgggctgcacaaagaagcagctgcccctgtgacattcacaaggcaatgtgaagaacaa 420  
 Qy 458 gacaccaaaggacacacagaagaagcacaagaacatcagcagcagcagcagcagcagc 517  
 Db 421 gacaccaaaggacacacagaagaagcacaagaacatcagcagcagcagcagcagcagc 480  
 Qy 518 aacaatgtcagtgcaagaagcttgctcctttagagccttgagcagcagcagcagcagc 577  
 Db 481 aacaatgtcagtgcaagaagcttgctcctttagagccttgagcagcagcagcagcagc 540  
 Qy 578 caattaacattctcagcagaagaagcagtgagcagcagcagcagcagcagcagcagcagc 637  
 Db 541 caattaacattctcagcagaagaagcagtgagcagcagcagcagcagcagcagcagcagc 600  
 Qy 638 caccctactctccactgtaccaccccttaataattcagctgtctcaaaagcagctt 697

Db 601 caccctactctccacatgctaccaccctaaatcattccagctgcctcaaaagatglt 660  
Oy 698 ttcaagatcattgtttgtgtgcctctctagtgtctctctctctgtcgaacttacc 757  
Db 661 ttcaagatcattgtttgtgtgcctctctagtgtctctctctctgtcgaacttacc 720  
Oy 758 ttgtgccctcccttaccagagcttagcttaactcctgaagaagctcaggaactgtag 817  
Db 721 ttgtgccctcccttaccagagcttagcttaactcctgaagaagctcaggaactgtag 780  
Oy 818 ctctcagctagtgtcatttaaccttaaatgcaatcaggaagtagcaacagaagtcac 877  
Db 781 ctctcagctagtgtcatttaaccttaaatgcaatcaggaagtagcaacagaagtcac 840  
Oy 878 taaatattttaaatgtc 895  
Db 841 taaatattttaaatgtc 858

## RESULT 5

AAZ65001  
ID AAZ65001 standard; cDNA; 870 BP.

XX AAZ65001;

DT 05-APR-2000 (first entry)

DE Membrane-bound protein PRO842 encoding cDNA.

KW Membrane-bound polypeptide; PRO polypeptide; IDL receptor; TIE ligand;

KV pharmaceutical; receptor immunoadhesin; gene mapping; ss.

XX Homo sapiens.

PM W09963088-A2.

PD 09-DEC-1999.

PF 02-JUN-1999; 99WO-US12252.

XX 02-JUN-1998; 98US-0087607.  
PR 02-JUN-1998; 98US-0087609.  
PR 02-JUN-1998; 98US-0087759.  
PR 03-JUN-1998; 98US-0087827.  
PR 04-JUN-1998; 98US-0088021.  
PR 04-JUN-1998; 98US-0088025.  
PR 04-JUN-1998; 98US-0088028.  
PR 04-JUN-1998; 98US-0088029.  
PR 04-JUN-1998; 98US-0088030.  
PR 04-JUN-1998; 98US-0088033.  
PR 04-JUN-1998; 98US-0088326.  
PR 05-JUN-1998; 98US-0088167.  
PR 05-JUN-1998; 98US-0088202.  
PR 05-JUN-1998; 98US-0088212.  
PR 05-JUN-1998; 98US-0088217.  
PR 09-JUN-1998; 98US-0088655.  
PR 10-JUN-1998; 98US-0088722.  
PR 10-JUN-1998; 98US-0088730.  
PR 10-JUN-1998; 98US-0088734.  
PR 10-JUN-1998; 98US-0088738.  
PR 10-JUN-1998; 98US-0088740.  
PR 10-JUN-1998; 98US-0088741.  
PR 10-JUN-1998; 98US-0088742.  
PR 10-JUN-1998; 98US-0088810.  
PR 10-JUN-1998; 98US-0088811.  
PR 10-JUN-1998; 98US-0088824.  
PR 10-JUN-1998; 98US-0088825.  
PR 10-JUN-1998; 98US-0088826.  
PR 11-JUN-1998; 98US-0088858.  
PR 11-JUN-1998; 98US-0088861.  
PR 11-JUN-1998; 98US-0088863.  
PR 11-JUN-1998; 98US-0088876.  
PR 12-JUN-1998; 98US-0089090.

PR 12-JUN-1998; 98US-0089105.  
PR 16-JUN-1998; 98US-0089440.  
PR 16-JUN-1998; 98US-0089512.  
PR 16-JUN-1998; 98US-0089514.  
PR 17-JUN-1998; 98US-0089532.  
PR 17-JUN-1998; 98US-0089538.  
PR 17-JUN-1998; 98US-0089598.  
PR 17-JUN-1998; 98US-0089599.  
PR 17-JUN-1998; 98US-0089600.  
PR 17-JUN-1998; 98US-0089609.  
PR 17-JUN-1998; 98US-0089653.  
PR 18-JUN-1998; 98US-0089801.  
PR 18-JUN-1998; 98US-0089807.  
PR 18-JUN-1998; 98US-0089908.  
PR 19-JUN-1998; 98US-0089947.  
PR 19-JUN-1998; 98US-0089948.  
PR 19-JUN-1998; 98US-0089952.  
PR 22-JUN-1998; 98US-0090246.  
PR 22-JUN-1998; 98US-0090252.  
PR 22-JUN-1998; 98US-0090254.  
PR 23-JUN-1998; 98US-0090349.  
PR 23-JUN-1998; 98US-0090355.  
PR 24-JUN-1998; 98US-0090429.  
PR 24-JUN-1998; 98US-0090431.  
PR 24-JUN-1998; 98US-0090435.  
PR 24-JUN-1998; 98US-0090444.  
PR 24-JUN-1998; 98US-0090445.  
PR 24-JUN-1998; 98US-0090461.  
PR 24-JUN-1998; 98US-0090472.  
PR 24-JUN-1998; 98US-0090535.  
PR 24-JUN-1998; 98US-0090538.  
PR 24-JUN-1998; 98US-0090540.  
PR 24-JUN-1998; 98US-0090557.  
PR 25-JUN-1998; 98US-0090676.  
PR 25-JUN-1998; 98US-0090678.  
PR 25-JUN-1998; 98US-0090688.  
PR 25-JUN-1998; 98US-0090690.  
PR 25-JUN-1998; 98US-0090691.  
PR 25-JUN-1998; 98US-0090694.  
PR 25-JUN-1998; 98US-0090695.  
PR 25-JUN-1998; 98US-0090696.  
PR 26-JUN-1998; 98US-0090862.  
PR 26-JUN-1998; 98US-0090863.  
PR 01-JUL-1998; 98US-0091358.  
PR 01-JUL-1998; 98US-0091360.  
PR 01-JUL-1998; 98US-0091544.  
PR 02-JUL-1998; 98US-0091478.  
PR 02-JUL-1998; 98US-0091486.  
PR 02-JUL-1998; 98US-0091519.  
PR 02-JUL-1998; 98US-0091626.  
PR 02-JUL-1998; 98US-0091628.  
PR 02-JUL-1998; 98US-0091633.  
PR 02-JUL-1998; 98US-0091646.  
PR 02-JUL-1998; 98US-0091673.  
PR 07-JUL-1998; 98US-0091978.  
PR 09-JUL-1998; 98US-0091982.  
PR 10-JUL-1998; 98US-0092182.  
PR 10-JUL-1998; 98US-0092472.  
PR 20-JUL-1998; 98US-0093339.  
PR 30-JUL-1998; 98US-0094651.  
PR 04-AUG-1998; 98US-0095282.  
PR 04-AUG-1998; 98US-0095285.  
PR 04-AUG-1998; 98US-0095301.  
PR 04-AUG-1998; 98US-0095302.  
PR 04-AUG-1998; 98US-0095318.  
PR 04-AUG-1998; 98US-0095321.  
PR 04-AUG-1998; 98US-0095325.  
PR 10-AUG-1998; 98US-0095916.  
PR 10-AUG-1998; 98US-0095929.  
PR 10-AUG-1998; 98US-0096102.  
PR 11-AUG-1998; 98US-0096143.  
PR 11-AUG-1998; 98US-0096146.  
PR 12-AUG-1998; 98US-0096329.  
PR 17-AUG-1998; 98US-0096757.

PR 17-AUG-1998; 98US-0096766.  
 PR 17-AUG-1998; 98US-0096768.  
 PR 17-AUG-1998; 98US-0096773.  
 PR 17-AUG-1998; 98US-0096791.  
 PR 17-AUG-1998; 98US-0096867.  
 PR 17-AUG-1998; 98US-0096891.  
 PR 17-AUG-1998; 98US-0096894.  
 PR 17-AUG-1998; 98US-0096895.  
 PR 17-AUG-1998; 98US-0096897.  
 PR 18-AUG-1998; 98US-0096949.  
 PR 18-AUG-1998; 98US-0096950.  
 PR 18-AUG-1998; 98US-0096950.  
 PR 18-AUG-1998; 98US-0096960.  
 PR 18-AUG-1998; 98US-0097022.  
 PR 19-AUG-1998; 98US-0097141.  
 PR 20-AUG-1998; 98US-0097218.  
 PR 26-AUG-1998; 98US-0097951.  
 PR 26-AUG-1998; 98US-0097952.  
 PR 26-AUG-1998; 98US-0097954.  
 PR 26-AUG-1998; 98US-0097954.  
 PR 26-AUG-1998; 98US-0097955.  
 PR 26-AUG-1998; 98US-0097971.  
 PR 26-AUG-1998; 98US-0097974.  
 PR 26-AUG-1998; 98US-0097978.  
 PR 26-AUG-1998; 98US-0097979.  
 PR 26-AUG-1998; 98US-0097986.  
 PR 26-AUG-1998; 98US-0098014.  
 PR 31-AUG-1998; 98US-0098525.  
 PR 16-SEP-1998; 98US-0100634.  
 PR 12-JAN-1999; 99US-0115565.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX  
 PI Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;  
 PI Wood WI, Yuan J;  
 DR WPI: 2000-072883/06.  
 DR P-SDS: AAY66668.  
 XX  
 PT Membrane-bound proteins and related nucleotide sequences -  
 XX  
 PS Claim 2; Fig 98; 822p; English.  
 XX  
 CC The invention provides membrane-bound PRO polypeptides and  
 CC polynucleotides encoding them. The PRO sequences of the invention were  
 CC identified based on extracellular domain homology screening. The PRO  
 CC sequences have homology with proteins including LDL receptors, TIE  
 CC ligands and various enzymes. The membrane-bound proteins and receptor  
 CC immunoadhesins, for instance, can be used as therapeutic agents. Receptor  
 CC receptor-ligand interactions. The membrane-bound proteins can also be  
 CC employed for screening of potential peptide or small molecule inhibitors  
 CC of the relevant receptor/ligand interaction. The PRO encoding sequences  
 CC are useful as hybridization probes, in chromosome and gene mapping and in  
 CC the generation of antisense RNA and DNA. PRO nucleic acid sequences  
 CC will also be useful for the preparation of PRO polypeptides, especially  
 CC by recombinant techniques.  
 XX  
 SQ Sequence 870 BP; 251 A; 237 C; 176 G; 206 T; 0 other;

Query Match 95.2%; Score 852; DB 21; Length 870;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 852; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 44 CTCGCGCTAATGGAAGCGCTGCGCTGGAGTAAGATAGACCAAGCGTGTATC 103  
 DB 1 CTCGCGCTAATGGAAGCGCTGCGCTGGAGTAAGATAGACCAAGCGTGTATC 60  
 QY 104 CTGACCTAGTATCCCGAGGATCAGGAGCTCCAGCAGGAGGAACCTCATATCT 163  
 DB 61 CTGACCTAGTATCCCGAGGATCAGGAGCTCCAGCAGGAGGAACCTCATATCT 120

QY 164 tcaagcaacttacagctgcacccagcaagttgcagatgaagtcttaattcttcctctcc 223  
 DB 121 tcaagcaacttacagctgcacccagcaagttgcagatgaagtcttaattcttcctctcc 180  
 QY 224 tgttcctgcacactatgtctgatgtccatgtctctccttaagcactgaatccgggtgcga 293  
 DB 181 tgttcctgcacactatgtctgatgtccatgtctctccttaagcactgaatccgggtgcga 240  
 QY 284 gaggccacagggacccagcagcagcagctctcagagatgtgtccacaggaagcgcgaagat 343  
 DB 241 gaggccacagggacccagcagcagcagcagctctcagagatgtgtccacaggaagcgcgaagat 300  
 QY 344 gtgagtgcacaaagtgtgtctctcagagcagcagcagcagcagcagcagcagcagcagcagc 403  
 DB 301 gtgagtgcacaaagtgtgtctctcagagcagcagcagcagcagcagcagcagcagcagcagc 360  
 QY 404 tgcacaagaagcagctgcctctgcagcagcagcagcagcagcagcagcagcagcagcagc 463  
 DB 361 tgcacaagaagcagctgcctctgcagcagcagcagcagcagcagcagcagcagcagcagc 420  
 QY 464 aaagcaccacagaagaagcacaacagcagcagcagcagcagcagcagcagcagcagcagc 523  
 DB 421 aaagcaccacagaagaagcacaacagcagcagcagcagcagcagcagcagcagcagcagc 480  
 QY 524 gtacgtcaagaagcgtgtgtct 583  
 DB 481 gtacgtcaagaagcgtgtgtct 540  
 QY 584 aacattctcagccacaagaagcagctgcagcagcagcagcagcagcagcagcagcagcagc 643  
 DB 541 aacattctcagccacaagaagcagctgcagcagcagcagcagcagcagcagcagcagcagc 600  
 QY 644 actctccacgttacccacccttaataatctccagctgtctctcctcaaaagcagcttcttcaa 703  
 DB 601 actctccacgttacccacccttaataatctccagctgtctctcctcaaaagcagcttcttcaa 660  
 QY 704 gatactttgtttgtgtctctctcagctgtctctctctctctctctctctctctctctctct 763  
 DB 661 gatactttgtttgtgtctctctcagctgtctctctctctctctctctctctctctctctct 720  
 QY 764 ctcccttaaccagcagcttagccttaattacgtgaagaattccaggaagaatcttagcttct 823  
 DB 721 ctcccttaaccagcagcttagccttaattacgtgaagaattccaggaagaatcttagcttct 780  
 QY 824 agctagtgcatttaaccttaattgaatcaggaagaagcagaagaagcagaagcagaata 883  
 DB 781 agctagtgcatttaaccttaattgaatcaggaagaagcagaagaagcagaagcagaata 840  
 QY 884 ttttaaatgtc 895  
 DB 841 ttttaaatgtc 852

RESULT 6  
 AAS4594  
 ID AAS4594 standard; cDNA; 870 BP.  
 XX  
 AC AAS4594;  
 XX  
 DT 18-DEC-2001 (first entry)  
 XX  
 DE Human DNA encoding PRO polypeptide sequence #70.  
 XX  
 KW PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep; ss;  
 KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;  
 KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;  
 KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder;  
 KW PCR primer.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200168848-A2.  
 XX

PD 20-SEP-2001.  
XX  
XX 28-FEB-2001; 2001MO-US06520.  
PR 01-MAR-2000; 2000MO-US05601.  
PR 02-MAR-2000; 2000MO-US05841.  
PR 03-MAR-2000; 2000US-187202P.  
PR 06-MAR-2000; 2000US-186968P.  
PR 14-MAR-2000; 2000US-189320P.  
PR 14-MAR-2000; 2000US-189328P.  
PR 15-MAR-2000; 2000MO-US06884.  
PR 21-MAR-2000; 2000US-190828P.  
PR 21-MAR-2000; 2000US-191007P.  
PR 21-MAR-2000; 2000US-191048P.  
PR 21-MAR-2000; 2000US-191314P.  
PR 28-MAR-2000; 2000US-192555P.  
PR 29-MAR-2000; 2000US-193032P.  
PR 29-MAR-2000; 2000US-193053P.  
PR 30-MAR-2000; 2000MO-US08439.  
PR 04-APR-2000; 2000US-194449P.  
PR 04-APR-2000; 2000US-194647P.  
PR 11-APR-2000; 2000US-195975P.  
PR 11-APR-2000; 2000US-196000P.  
PR 11-APR-2000; 2000US-196187P.  
PR 11-APR-2000; 2000US-196690P.  
PR 11-APR-2000; 2000US-196820P.  
PR 18-APR-2000; 2000US-198121P.  
PR 18-APR-2000; 2000US-198585P.  
PR 25-APR-2000; 2000US-199397P.  
PR 25-APR-2000; 2000US-199550P.  
PR 03-MAY-2000; 2000US-199654P.  
PR 17-MAY-2000; 2000MO-US13705.  
PR 22-MAY-2000; 2000MO-US14042.  
PR 30-MAY-2000; 2000MO-US14941.  
PR 02-JUN-2000; 2000MO-US15264.  
PR 05-JUN-2000; 2000US-209832P.  
PR 28-JUL-2000; 2000MO-US20710.  
PR 22-AUG-2000; 2000US-0644848.  
PR 24-AUG-2000; 2000MO-US23328.  
PR 08-NOV-2000; 2000MO-US30952.  
PR 01-DEC-2000; 2000MO-US32678.  
PR 20-DEC-2000; 2000MO-US34956.  
XX  
XX (GETH ) GENENTECH INC.  
XX  
PI Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;  
PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;  
XX  
XX WPI; 2001-602746/68.  
DR P-PSDB: AAN29093.  
XX  
XX Novel nucleic acids encoding PRO polypeptides, used to diagnose the  
PT presence of tumours, such as prostate and breast tumours, in mammals and  
PT to screen for modulators of the compounds -  
XX  
XX Claim 2; Fig 139; 774pp; English.  
XX  
XX Sequences AAS45925-AAS46231 represent DNA molecules encoding and PCR  
CC primers for PRO polypeptides of the invention. The sequences of the  
CC invention can be used to detect the presence of a tumour in a mammal by  
CC comparing the level of expression of a PRO polypeptide in a test sample  
CC of cells from the animal and a control sample of normal cells, whereby a  
CC higher level of expression in the test sample indicates the presence of a  
CC tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep,  
CC pigs, goats and rabbits but are preferably human. The polypeptides can be  
CC used to stimulate tumour necrosis factor (TNF) alpha release from human  
CC blood, when contacted with it. A specific polypeptide can be used to  
CC stimulate the proliferation or differentiation of chondrocyte cells. The  
CC PRO proteins can be used to determine the presence of tumours and also  
CC susceptibility to tumour development, particularly adrenal, lung, colon,  
CC breast, prostate, rectal, cervical, or liver tumours, in mammalian  
CC subjects. The oligonucleotide probes specific for the PRO nucleic acids

CC can be used for genetic analysis of individuals with genetic disorders.  
XX  
XX Sequence 870 BP; 251 A; 237 C; 176 G; 206 T; 0 other;  
SQ

## Query Match

95.2%; Score 852; DB 22; Length 870;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 852; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 ctgcgccctcaaatggaagcgtcgccctggaactaaagcacaagcagcgtgagtc 103  
|||||  
DB 1 ctgcgccctcaaatggaagcgtcgccctggaactaaagcacaagcagcgtgagtc 60  
QY 104 ctgacgtgagtcattcccccaggagatcaggagcctccagcagggaaccccttcattatct 163  
|||||  
DB 61 ctgacgtgagtcattcccccaggagatcaggagcctccagcagggaaccccttcattatct 120  
QY 164 tcaagcaactacagctgacccagcagcttgatgataagttctaatctctccctcc 223  
|||||  
DB 121 tcaagcaactacagctgacccagcagcttgatgataagttctaatctctccctcc 180  
QY 224 tcttgctgcactaatgtgtatgtlccatggtctctagcagcctgaaatccaggggtcgca 283  
|||||  
DB 181 tcttgctgcactaatgtgtatgtlccatggtctctagcagcctgaaatccaggggtcgca 240  
QY 284 gaggccacagggagccgagccagcgtcttaaggatggtctccagggagccgccaagaat 343  
|||||  
DB 241 gaggccacagggagccgagccagcgtcttaaggatggtctccagggagccgccaagaat 300  
QY 344 gtgagtgcaagaatctgtctctgagagcccgagaaataatcagaaagtgtctgggc 403  
|||||  
DB 301 gtgagtgcaagaatctgtctctgagagcccgagaaataatcagaaagtgtctgggc 360  
QY 404 tgcacaaagaagcagtgccctgtgatcatttcaaggcaatgtgaaagaacaagacacc 463  
|||||  
DB 361 tgcacaaagaagcagtgccctgtgatcatttcaaggcaatgtgaaagaacaagacacc 420  
QY 464 aagagccacacagaaagcacaacaaagcattccagaaagcgtccagcaattctcaacaat 523  
|||||  
DB 421 aagagccacacagaaagcacaacaaagcattccagaaagcgtccagcaattctcaacaat 480  
QY 524 gtcagctaaagaagcttgcctctgctcttgtagagctctgagcgtccactctccaatta 583  
|||||  
DB 481 gtcagctaaagaagcttgcctctgctcttgtagagctctgagcgtccactctccaatta 540  
QY 584 aacattctcagccaaagaagcagtgacacactaccagaacactctctctccaccctc 643  
|||||  
DB 541 aacattctcagccaaagaagcagtgacacactaccagaacactctctctccaccctc 600  
QY 644 actctccacatgacccacccccaataatcattccagtgctctcaaaaagcatgttttcaa 703  
|||||  
DB 601 actctccacatgacccacccccaataatcattccagtgctctcaaaaagcatgttttcaa 660  
QY 704 gatcatttgttctgtctctctctagtgctctctctctctctctctctctctctctctct 763  
|||||  
DB 661 gatcatttgttctgtct 720  
QY 764 ctcccttaccagcgtctggttaattacctcgaagaagattccaagaactgtagcttct 823  
|||||  
DB 721 ctcccttaccagcgtctggttaattacctcgaagaagattccaagaactgtagcttct 780  
QY 824 agctagtgcatcttaacctaagtcacatcaggaagtgcaaaagcagtgcaataata 883  
|||||  
DB 781 agctagtgcatcttaacctaagtcacatcaggaagtgcaaaagcagtgcaataata 840  
QY 884 ttcttaacgtc 895  
|||||  
DB 841 ttcttaacgtc 852  
RESULT 7  
AAF92070  
ID AAF92070 standard; cDNA; 870 BP.

XX AA92070;  
AC  
XX  
DT 15-MAY-2001 (first entry)  
XX  
DE Human PRO842 cDNA.  
XX  
KW Human; PRO protein; mapping; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200116318-A2.  
XX  
PD 08-MAR-2001.  
XX  
PE 24-AUG-2000; 2000WO-US23328.  
XX  
PR 01-SEP-1999; 99WO-US20111.  
PR 15-SEP-1999; 99WO-US21080.  
PR 07-DEC-1999; 99US-0169495.  
PR 09-DEC-1999; 99US-0170262.  
PR 11-JAN-2000; 2000US-0175481.  
PR 18-FEB-2000; 2000WO-US04341.  
PR 18-FEB-2000; 2000WO-US04342.  
PR 22-FEB-2000; 2000WO-US04414.  
PR 01-MAR-2000; 2000WO-US05601.  
PR 03-MAR-2000; 2000US-0187202.  
PR 23-APR-2000; 2000US-0193937.  
PR 22-MAY-2000; 2000WO-US14042.  
PR 05-JUN-2000; 2000US-0209832.  
XX  
PA (GETH ) GENENTECH INC.  
XX  
PI Epton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;  
PI Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;  
XX  
XX WPI; 2001-183260/18.  
DR P-PSDB; AAB87538.  
XX  
XX  
XX Eighty four nucleic acids encoding PRO polypeptides, useful in  
PT molecular biology, including use as hybridization probes, and in  
PT chromosome and gene mapping.  
XX  
XX Claim 2; Fig 25; 278pp; English.  
XX  
XX The present sequence is the coding sequence for a human PRO polypeptide  
CC (secreted and transmembrane). The PRO protein, and PRO agonists, PRO  
CC antagonists or anti-PRO antibodies are useful for preparation of a  
CC medicament useful in the treatment of a condition which is responsive to  
CC the PRO protein, agonists, antagonists or anti-PRO antibodies. The PRO  
CC protein may also be employed as molecular weight markers for protein  
CC electrophoresis. The PRO coding sequence has applications in molecular  
CC biology, including use as hybridisation probes, and in chromosome and  
CC gene mapping.  
XX  
XX  
SQ Sequence 870 BP; 251 A; 237 C; 176 G; 206 T; 0 other;

Query Match 95.2%; Score 852; DB 22; Length 870;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 852; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 44 ctgcacctaaatggaagcgtgctggaactaaagacacacacagcgtgagatc 103  
DB 1 ctgcacctaaatggaagcgtgctggaactaaagacacacacagcgtgagatc 60  
QY 104 ctgacctgagatccccaaggaatcagaagcctccagcaggaagaaactcattatct 163  
DB 61 ctgacctgagatccccaaggaatcagaagcctccagcaggaagaaactcattatct 120  
QY 164 tcaagaacttaacgacgtgacagagatgagatgaagttcattctctccctcc 223  
DB 121 tcaagaacttaacgacgtgacagagatgagatgaagttcattctctccctcc 180

QY 224 tgtgtcgtccactaatgctatgctatcattgctctctctagcaagcctgaatccagggtctgc 283  
DB 181 tgtgtcgtccactaatgctatgctatcattgctctctctagcaagcctgaatccagggtctgc 240  
QY 284 gaggccacaggaagcaggaagcagcgtctctaggaatggtctccaggaagcgcgaagaat 343  
DB 241 gaggccacaggaagcaggaagcagcgtctctaggaatggtctccaggaagcgcgaagaat 300  
QY 344 gtgagtgcaaatgtgtctctgagagcccggaagaagaatctatcagtgctgtggc 403  
DB 301 gtgagtgcaaatgtgtctctgagagcccggaagaagaatctatcagtgctgtggc 360  
QY 404 tgcacaagaagcagtgccctctgatactatcattcaagggaatgtgaagaacaagacacc 463  
DB 361 tgcacaagaagcagtgccctctgatactatcattcaagggaatgtgaagaacaagacacc 420  
QY 464 aaagccacacagaaagcacaagcattccagagcctgcaggaattctcaaacat 523  
DB 421 aaagccacacagaaagcacaagcattccagagcctgcaggaattctcaaacat 480  
QY 524 gtcaagtaagaagcctgtctctgctcttgtagagcctctgagccacactcacaata 583  
DB 481 gtcaagtaagaagcctgtctctgctcttgtagagcctctgagccacactcacaata 540  
QY 584 aacattctcagcacaagaagcagtgagcagcactaccagacactctcttccacc 643  
DB 541 aacattctcagcacaagaagcagtgagcagcactaccagacactctcttccacc 600  
QY 644 actctccactgtaccacacccctaatactcaccagtcctcacaagaagcatttca 703  
DB 601 actctccactgtaccacacccctaatactcaccagtcctcacaagaagcatttca 660  
QY 704 gatcattgt 763  
DB 661 gatcattgt 720  
QY 764 ctcccttaccacagcgttagccttaattacctgaagaatccaggaatctgtgtgtgt 823  
DB 721 ctcccttaccacagcgttagccttaattacctgaagaatccaggaatctgtgtgtgt 780  
QY 824 agctagtgatcatttaacacttaaatgcaatcaggaagaatgacaaagatcaataa 883  
DB 781 agctagtgatcatttaacacttaaatgcaatcaggaagaatgacaaagatcaataa 840  
QY 884 tttttaaatgtc 895  
DB 841 tttttaaatgtc 852

RESULT 8  
AA944147  
ID AA944147 standard; cDNA; 870 BP.  
XX  
AC AA944147;  
XX  
DT 02-APR-2001 (first entry)  
XX  
DE Human PRO842 (U00473) nucleotide sequence SEQ ID NO:164.  
XX  
KW Human; secreted and transmembrane protein; PRO; cytosolic;  
KW cell death; cancer; chromosomal mapping; gene mapping; tissue typing;  
KW diagnostic assay; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO200073454-A1.  
XX  
XX  
XX 07-DEC-2000.  
XX  
XX 30-MAR-2000; 2000WO-US08439.  
XX  
XX  
XX 02-JUN-1999; 99WO-US12252.  
PR

PR 23-JUN-1999; 99US-0141037.  
PR 07-JUL-1999; 99US-0143048.  
PR 20-JUL-1999; 99US-0144758.  
PR 26-JUL-1999; 99US-0145698.  
PR 28-JUL-1999; 99US-0146222.  
PR 17-AUG-1999; 99US-0149396.  
PR 15-SEP-1999; 99WO-US21090.  
PR 15-SEP-1999; 99WO-US21547.  
PR 08-OCT-1999; 99US-0158663.  
PR 30-NOV-1999; 99WO-US28313.  
PR 01-DEC-1999; 99WO-US28301.  
PR 16-DEC-1999; 99WO-US30095.  
PR 20-DEC-1999; 99WO-US30911.  
PR 05-JAN-2000; 2000WO-US00219.  
PR 06-JAN-2000; 2000WO-US00376.  
PR 11-FEB-2000; 2000WO-US03565.  
PR 18-FEB-2000; 2000WO-US04341.  
PR 22-FEB-2000; 2000WO-US04414.  
PR 24-FEB-2000; 2000WO-US04914.  
PR 24-FEB-2000; 2000WO-US05004.  
PR 02-MAR-2000; 2000WO-US05884.  
PR 15-MAR-2000; 2000WO-US06884.  
PR 20-MAR-2000; 2000WO-US07377.  
XX  
XX (GETH ) GENENTECH INC.  
PA  
XX  
PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;  
PI Ferrara N, Fong S, Gerber H, Gerritsen ME, Goddard A, Godowski PJ;  
PI Grimaldi CJ, Gurney AL, Kijavlin IJ, Napier MA, Pan J, Faoni NF;  
PI Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Wood WI;  
PI Zhang Z;  
XX  
XX WPI: 2001-032160/04.  
DR P-PSDB: AAB65191.  
XX  
XX PRO polynucleotides used to produce polypeptides used to target  
PT bioactive molecules such as toxins, radiolabels or antibodies, to  
PT specific cells, to cause targeted cell death -  
XX  
XX  
PS Claim 2; Fig 98; 935pp; English.  
XX  
XX The present invention describes human secreted and transmembrane PRO  
CC proteins. The PRO proteins have cytostatic activity. The PRO proteins  
CC can be used for targeted delivery of bioactive molecules, such as  
CC toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide  
CC sequences, and their fragments, can be used as hybridisation probes, in  
CC chromosomal and gene mapping, and in the generation of anti-sense RNA  
CC and DNA. They may also be used to produce transgenic animals which are  
CC used to develop and screen therapeutically useful reagents. The PRO  
CC nucleotide and protein sequence can be used for tissue typing and in  
CC treating cancer. Anti-PRO antibodies can be used in diagnostic assays.  
CC AAF44270 to AAF44470 represent PCR primers and hybridisation probes used  
CC in the isolation of human PRO sequences. AAF44087 to AAF44269 and  
CC AAB65154 to AAB65300 represent human PRO polynucleotide and protein  
CC sequences given in the exemplification of the present invention.  
XX  
XX  
SQ Sequence 870 BP; 251 A; 237 C; 176 G; 206 T; 0 other;

Query Match 95.2%; Score 852; DB 22; Length 870;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 852; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 ctgcgcctcaaatgggaacgctgacctgggactaagaacacacacagctgagtc 103  
DB 1 ctgcgcctcaaatgggaacgctgacctgggactaagaacacacacagctgagtc 60  
QY 104 ctgacctgagtcattcccgaggaatcagaagcctcagcagggagaaccttcattatctt 163  
DB 61 ctgacctgagtcattcccgaggaatcagaagcctcagcagggagaaccttcattatctt 120  
QY 164 tcaagaaccttaacagctcagccagacagctgcatgaaattcattcttcctcctcc 223

DB 121 tcaagaaccttaacagctcagccagacagctgcatgaaattcattcttcctcctcc 180  
QY 224 tcttgctgccactaatgtcgtatccatgctctctcagcagccttgatccagggtgcga 283  
DB 181 tcttgctgccactaatgtcgtatccatgctctctcagcagccttgatccagggtgcga 240  
QY 284 gaggccacagagcagcagcagcagcctctcagagatctgctcaggaagcagcagaagt 343  
DB 241 gaggccacagagcagcagcagcagcctctcagagatctgctcaggaagcagcagaagt 300  
QY 344 gtgagtgcaaaagtgtgtctctgagagcccgagaagaataattcattgacagtgtctgggc 403  
DB 301 gtgagtgcaaaagtgtgtctctgagagcccgagaagaataattcattgacagtgtctgggc 360  
QY 404 tgcgaagaagcagtgccctgtgatactttcaaggcagatgtgaaagaaacaaagaccc 463  
DB 361 tgcgaagaagcagtgccctgtgatactttcaaggcagatgtgaaagaaacaaagaccc 420  
QY 464 aagagcaccacagaaagcacaacagcattccagagcctgcagcaattctcacaacat 523  
DB 421 aagagcaccacagaaagcacaacagcattccagagcctgcagcaattctcacaacat 480  
QY 524 gtcaagtaagaagcttgctctgctcttgtagagcctctgagcgcacctctccaatta 583  
DB 481 gtcaagtaagaagcttgctctgctcttgtagagcctctgagcgcacctctccaatta 540  
QY 584 aacattctcagcacaagaagacagtgagcaccctccagacactctctctccacacc 643  
DB 541 aacattctcagcacaagaagacagtgagcaccctccagacactctctctccacacc 600  
QY 644 actctccacgtaccacccctaaatcattcagtgctctcaaaaagcatgttttcaa 703  
DB 601 actctccacgtaccacccctaaatcattcagtgctctcaaaaagcatgttttcaa 660  
QY 704 gatcatttgttgtgtctctctcagtgctctctctctctctcagcagctttagcctgtgc 763  
DB 661 gatcatttgttgtgtctctctcagtgctctctctctctctcagcagctttagcctgtgc 720  
QY 764 ctcccttaaccagcttagcttaattaccctgaaagatctcaggaatctgtagcttct 823  
DB 721 ctcccttaaccagcttagcttaattaccctgaaagatctcaggaatctgtagcttct 780  
QY 824 agctagtgcttaaccttaattgcacatcagagaaagtagcaacagagtaataata 883  
DB 781 agctagtgcttaaccttaattgcacatcagagaaagtagcaacagagtaataata 840  
QY 884 ttcttaaatgtc 895  
DB 841 ttcttaaatgtc 852

RESULT 9  
ABA09257  
ID ABA09257 standard; cDNA; 1148 BP.  
XX  
AC ABA09257;  
XX  
DT 11-JAN-2002 (first entry)  
XX  
DE Human TGC-440 homologue-encoding cDNA, SEQ ID NO:1033.  
XX  
XX Human: cytokine; cell proliferation; tissue growth; immunomodulator; growth factor;  
KW haematopoiesis regulation; tumour; cancer; lymphoid cell disorder; asthma; arthritis;  
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;  
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
KW chronic inflammatory condition; proliferative retinopathy;  
KW atherosclerosis; coronary heart disease; arterial ischaemia;  
KW bone disorder; osteoporosis; vascular growth disorder;  
KW tissue regeneration; wound healing; infection; immune disorder;  
KW cell culture; drug screening; gene therapy; antiinflammatory;  
KW antasthmatic; antiarthritis; haemostatic; antiarteriosclerotic;  
KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;







```
PD 16-MAR-2000.
XX
XX 02-SEP-1999; 99WO-JP04765.
XX
PR 03-SEP-1998; 98JP-0250108.
XX
XX (TAKE ) TAKEDA CHEM IND LTD.
XX
XX Itoh Y, Ogi K, Tanaka H, Kitada C;
XX WPI: 2000-256978/22.
DR P-PSDB: AAY82453.
XX
PT Secretory protein TGC440, antibodies to it and compounds promoting or
PT inhibiting its activity for diagnosis and treatment of diseases of the
XX immune system, lung, kidney, liver and intestinal system
XX
XX Disclosure; Fig 1; 86pp; Japanese.
XX
XX The present sequence encodes a human secretory protein designated
CC TGC-440. TGC-440 has antiinflammatory, nephrotropic, immunomodulatory,
CC virucide, hepatotropic, antilasthmatic and antibacterial activities,
CC and can be used in vaccines. TGC-440 and the polynucleotide sequence
CC encoding it can be used to treat, prevent and diagnose immunological,
CC lung, liver, kidney or gastrointestinal disorders and infectious
CC diseases, such as hepatitis, nephritis, influenza, asthma, pneumonia,
CC pulmonary hypertension, and Helicobacter pylori infection. An antibody
CC immunospecific for TGC-440 is also useful in the above treatment and
CC diagnosis, and also for quantifying the amount of TGC-440 in a liquid
CC specimen.
XX
XX Sequence 849 BP; 250 A; 219 C; 178 G; 202 T; 0 other;
SQ
Query Match 66.3%; Score 593; DB 21; Length 849;
Best Local Similarity 100.0%; Pred. No. 8.3e-291;
Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ctaatctgtacgtaacagacagacagctacactcactgttcttcgcacctaaatggga 60
DB 25 ctatctcgttaacgtaacagacagacagcgtcacctcactcgttcttcgcacctaaatggga 84
QY 61 acgctgacctggagactaaagacatagaccaccagcgtgagtatctcctcctggtatccccc 120
DB 85 acgctgacctggagactaaagacatagaccaccagcgtgagtatctcctcctggtatccccc 144
QY 121 caggagatcaggagcctccagcagaggaaacctccatatacttccaagaacttaagct 180
DB 145 caggagatcaggagcctccagcagaggaaacctccatatacttccaagaacttaagct 204
QY 181 gaccgacaggttgcagatgaagaattcctaactcttcctcctcctcctgttgcctcaactatg 240
DB 205 gaccgacaggttgcagatgaagaattcctaactcttcctcctcctcctgttgcctcaactatg 264
QY 241 ctgagtgcacatgtctctcagcagcctgaaatccaggggtgcgcagaggccacagggagccga 300
DB 265 ctgagtgcacatgtctctcagcagcctgaaatccaggggtgcgcagaggccacagggagccga 324
QY 301 ggcacggccttcaagagatggtctccaaggaaggcgccaaagaatgtgagtcaagaattgg 360
DB 325 ggcacggccttcaagagatggtctccaaggaaggcgccaaagaatgtgagtcaagaattgg 384
QY 361 ttccctgagagcccgagagaagaattcagacagtgtctgggcttccaaagaagcagatgc 420
DB 385 ttccctgagagcccgagagaagaattcagacagtgtctgggcttccaaagaagcagatgc 444
QY 421 cccctgtgacatctcaagggaatgtgaagaacaagacaccaaaggccacacagaatg 480
DB 445 cccctgtgacatctcaagggaatgtgaagaacaagacaccaaaggccacacagaatg 504
QY 481 ccaacaagacatccagagcctgcgcagcaatttccaacaagtgcagctagaagcttt 540
DB 505 ccaacaagacatccagagcctgcgcagcaatttccaacaagtgcagctagaagcttt 564
```

```
QY 541 gctctgcttctgagagcctctgagcgcaccactctccaatctcaatctctca 593
DB 565 gctctgcttctgagagcctctgagcgcaccactctccaatctcaatctctca 617
RESULT 12
ID AAF68195 standard; cDNA; 533 BP.
XX
XX AAF68195;
XX
XX 12-APR-2001 (first entry)
DE Human lung tumour protein related nucleotide sequence SEQ ID NO:113.
XX
XX Human; lung cancer; lung tumour; lung tumour protein; gene therapy;
KW lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;
KW cytostatic; antisense inhibition; ss.
XX
XX Homo sapiens.
XX
XX WO200100828-A2.
XX
XX 04-JAN-2001.
XX
XX 30-JUN-2000; 2000WO-US18061.
XX
XX 30-JUN-1999; 99US-0346492.
XX
XX 15-OCT-1999; 99US-0419356.
XX
XX 17-DEC-1999; 99US-0466867.
XX
XX 30-DEC-1999; 99US-0476300.
XX
XX 06-MAR-2000; 2000US-0519642.
XX
XX 22-MAR-2000; 2000US-0533077.
XX
XX 10-APR-2000; 2000US-0546259.
XX
XX 27-APR-2000; 2000US-0560406.
XX
XX 05-JUN-2000; 2000US-0589184.
XX
XX (CORI-) CORIXA CORP.
XX
XX Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;
XX Retter MW, Mannion J;
XX WPI: 2001-071488/08.
XX
XX Claim 4; Page 186; 436pp; English.
XX
XX The present invention describes immunogenic portions of lung tumour-
XX associated proteins (I) and the nucleic acids (NAs) that encode them.
XX (I) have cytostatic activity and can be used in gene therapy, antisense
XX inhibition and in vaccines. The NAs and the lung tumour-associated
XX proteins they encode may be used in the prevention, treatment and
XX diagnosis of diseases associated with their inappropriate expression,
XX especially lung cancers. For example, the NAs may be administered to
XX treat diseases by rectifying mutations or deletions in a patient's genome
XX that affect the activity of the protein by expressing inactive proteins
XX or to supplement the patient's own production of (I). Additionally, the
XX NAs may be used to produce the lung-tumour associated protein, according
XX to standard recombinant DNA methodology. Conversely, antisense NA
XX molecules may be administered to down regulate protein expression by
XX binding with the cells own genes and preventing their expression. The NA
XX and complementary sequences may also be used as DNA probes in diagnostic
XX assays to detect and quantitate the presence of similar NA sequences in
XX cancer. The (I) may be used as antigens in the production of antibodies
XX and in assays to identify modulators (agonists and antagonists) of the
XX expression and activity of the protein. AAF68083 to AAF68878 and
XX AAF76848 to AAF76878 represent human lung tumour protein related
XX nucleotide and protein sequences which are used in the exemplification
XX of the present invention.
```



KM forensic; gene therapy; chromosome mapping; signal peptide; prostate;  
 KM upstream regulatory sequence; cytokine activity; cell proliferation;  
 KM differentiation; haematopoiesis regulation; tissue growth regulation;  
 KM reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
 KM thrombolytic; anti-inflammatory; tumour inhibition; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9906550-A2.  
 XX  
 PD 11-FEB-1999.  
 XX  
 PF 31-JUL-1998; 98WO-1B01232.  
 XX  
 PR 01-AUG-1997; 97US-0905144.  
 XX  
 PA (GEST ) GENSET.  
 XX  
 PI Duclert A, Dumas Milne Edwards J, Lacroix B;  
 DR WPI; 1999-153780/13.  
 DR P-PSDB; AAY11732.  
 XX  
 PT New isolated prostate-derived nucleic acids - used to develop  
 PT products which may have cytokine, immune regulatory, haematopoiesis  
 PT regulating, anti-inflammatory or tumour inhibition activity  
 XX  
 PS Claim 1; Page 185; 675pp; English.

CC AAX40438 to AAX40715 represent 5' expressed sequence tags (ESTs) for  
 CC human secreted proteins expressed in prostate, and encode the proteins  
 CC given in AAY11716 to AAY11993 respectively. The proteins given represent  
 CC the signal peptide and an N-terminal fragment of a secreted protein. The  
 CC nucleic acid sequences can be used for producing secreted human gene  
 CC products. They can also be used to develop products for diagnosis and  
 CC therapy. The proteins obtained may have cytokine activity, cell  
 CC proliferation and differentiation activity, haematopoiesis regulating  
 CC activity, tissue growth regulating activity, reproductive hormone  
 CC regulating activity, chemotactic/chemokinetic activity, haemostatic and  
 CC thrombolytic activity, receptor/ligand activity, anti-inflammatory  
 CC activity, tumour inhibition activity or other activities. The products  
 CC can be used in forensic, gene therapy and chromosome mapping procedures.  
 CC The sequences can also be used for obtaining corresponding promoter  
 CC sequences. The nucleic acids encoding the signal peptides can be used for  
 CC directing extracellular secretion of a polypeptide or the insertion of a  
 CC polypeptide into a membrane, or importing a polypeptide into a cell.  
 CC  
 SQ Sequence 365 BP; 91 A; 100 C; 96 G; 78 T; 0 other;

Query Match 37.4%; Score 335; DB 20; Length 365;  
 Best Local Similarity 100.0%; Pred. No. 8e-160;  
 Matches 335; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 68 cctggagcctaagacatgaccacgagctgagatctgagcctgagatcattcccaaggat 127  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 31 cctggagcctaagacatgaccacgagctgagatcattcccaaggatcctcccaaggat 90  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 128 caggagcctcagcaggaaccttcattatattctcaagaactaacagtcgaccca 187  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 91 caggagcctcagcaggaaccttcattatattctcaagaactaacagtcgaccca 150  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 188 cagttcgatgaagtcttaattcttccctcctctgttgcctgacactaagtctgatt 247  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 151 cagttcgatgaagtcttaattcttccctcctcctgttgcctgacactaagtctgatt 210  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 248 ccatgtcctctagcagcctgaatccaggggtgcagagccacagggccagggccagc 307  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 211 cctatgtcctctagcagcctgaatccaggggtgcagagccacagggccagggccagc 270  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 308 cttctagagagatggtctccaggaagcgccaaagaatgtagtcgcaaaagttgttcccta 367  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 271 cttctagagagatggtctccaggaagcgccaaagaatgtagtcgcaaaagttgttcccta 330

QY 368 gagccccgagaagaattcatgacagtgctcgg 402  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 Db 331 gagccccgagaagaattcatgacagtgctcgg 365  
 RESULT 15  
 AAF68125  
 ID AAF68125 standard; cDNA; 312 BP.  
 XX  
 AC AAF68125;  
 XX  
 DT 12-APR-2001 (first entry)  
 XX  
 DE Human lung tumour protein related nucleotide sequence SEQ ID NO:43.  
 XX  
 KW Human; lung cancer; lung tumour; lung tumour protein; gene therapy;  
 KW lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;  
 KW cytostatic; antisense inhibition; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200100828-A2.  
 XX  
 PD 04-JAN-2001.  
 XX  
 PF 30-JUN-2000; 2000WO-US18061.  
 XX  
 PR 30-JUN-1999; 99US-0346492.  
 PR 15-OCT-1999; 99US-0419356.  
 PR 17-DEC-1999; 99US-0466867.  
 PR 30-DEC-1999; 99US-0476300.  
 PR 06-MAR-2000; 2000US-0519642.  
 PR 12-MAR-2000; 2000US-0533077.  
 PR 10-APR-2000; 2000US-0546259.  
 PR 27-APR-2000; 2000US-0560406.  
 PR 05-JUN-2000; 2000US-0589184.  
 XX  
 PA (COR-) CORIXA CORP.  
 XX  
 PI Wang T, Bangur GS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;  
 PI Retter MW, Mannion J;  
 XX  
 DR WPI; 2001-071488/08.  
 XX  
 PT Lung tumor-associated proteins and the nucleic acids that encode them,  
 PT useful for preventing, diagnosing and treating lung cancer -  
 XX  
 PS Claim 4; Page 166; 436pp; English.

The present invention describes immunogenic portions of lung tumour-associated proteins (I) and the nucleic acids (NAs) that encode them. (I) have cytostatic activity and can be used in gene therapy, antisense inhibition and in vaccines. The NAs and the lung tumour-associated proteins they encode may be used in the prevention, treatment and diagnosis of diseases associated with their inappropriate expression, especially lung cancers. For example, the NAs may be administered to treat diseases by rectifying mutations or deletions in a patient's genome that affect the activity of the protein by expressing inactive proteins or to supplement the patient's own production of (I). Additionally, the NAs may be used to produce the lung-tumour associated protein, according to standard recombinant DNA methodology. Conversely, antisense NA molecules may be administered to down regulate protein expression by binding with the cells own genes and preventing their expression. The NA and complementary sequences may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar NA sequences in samples. The (I) may be used as antigens in the production of antibodies and in assays to identify modulators (agonists and antagonists) of the expression and activity of the protein. AAF68083 to AAF68878 and AAF76848 to AAF76878 represent human lung tumour protein related nucleotide and protein sequences which are used in the exemplification of the present invention.

XX Sequence 312 BP; 100 A; 82 C; 69 G; 61 T; 0 other;

Query Match 34.9%; Score 312; DB 22; Length 312;  
 Best Local Similarity 100.0%; Pred. No. 3.8e-148;  
 Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 cagggaaggcgcccaagaatgtgagtgcaagaatgtgtctctgagagcccgagaagaaaa 384  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1 cagggaaggcgcccaagaatgtgagtgcaagaatgtgtctctgagagcccgagaagaaaa 60  
 QY 385 ttcatgacagtgctgtggctgccaagaagcagtgccctgtgatcatlccaaggccaat 444  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 61 ttcatgacagtgctgtggctgccaagaagcagtgccctgtgatcatlccaaggccaat 120  
 QY 445 gtgaagaataaacagacccaagaaggcacaagaagaacccaagaatccagagcctgc 504  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 121 gtgaagaataaacagacccaagaaggcacaagaagaacccaagaatccagagcctgc 180  
 QY 505 cagcaatttctcaacaatgtcagctagaagaagcttgcctgtgcttgtagaagctctga 564  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 181 cagcaatttctcaacaatgtcagctagaagaagcttgcctgtgcttgtagaagctctga 240  
 QY 565 gcgcccactcttccaaatlaacattctcagccaagaagaacagtgagcacacctacagac 624  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 241 gcgcccactcttccaaatlaacattctcagccaagaagaacagtgagcacacctacagac 300  
 QY 625 actcttctcttc 636  
 ||||||||||||  
 Db 301 actcttctcttc 312

Search completed: September 16, 2002, 06:31:48  
 Job time: 7656 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 02:53:46 ; Search time 3785.3 Seconds  
(without alignments)  
1741.436 Million cell updates/sec

Title: US-09-700-770-4

Perfect score: 315  
Sequence: 1 taacactgactacattt.....tgtgtattactactatgc 315

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_hcg.\*  
3: gb\_in.\*  
4: gb\_on.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_hcg\_hum.\*  
31: em\_hcg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query Match	Length	DB ID	Description
------------	-------------	--------	-------	-------------

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	AL356798/c	Human DNA sequence from clone Rpl1-467B11 on chromosome 9, complete sequence.	AL356798	1	GI:15795413	human.	1	Babbage,A.	Direct Submission	Submitted (26-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk	On Sep 27, 2001 this sequence version replaced gi:14455883. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission
2	40.4	12.8	226399	2	AF425794	Human DNA	2	AC025794	Mus muscu		
3	40	12.7	125020	2	AF429315	Mus muscu	2	AF429315	Mus muscu		
4	39.8	12.6	201794	2	CNS0180W	Human chr	2	AL109759	Human chr		
5	39.6	12.6	62871	2	AC102444	Mus muscu	2	AF318573	Bovine he		
6	39.6	12.5	54700	3	CEY116A8	Continuation (2 of	3	CEY116A8_1	Continuation (2 of		
7	39.4	12.5	110000	2	CEY113B8_1	Continuation (3 of	2	CEY113B8_2	Continuation (3 of		
8	39.4	12.5	110000	2	CEY113B8_2	Continuation (3 of	2	CEY113B8_1	Continuation (3 of		
9	39.4	12.5	110000	2	CEY113B8_1	Continuation (3 of	2	CEY113B8_2	Continuation (3 of		
10	39.4	12.5	110000	2	CEY113B8_2	Continuation (3 of	2	CEY113B8_1	Continuation (3 of		
11	39.4	12.4	150355	2	HS364H10	Human trolg	2	AL078603	Human trolg		
12	39	12.4	168043	2	AC011935	Human sapi	2	AC011935	Human sapi		
13	39	12.4	174576	2	AC013287	Human sapi	2	AC013287	Human sapi		
14	39	12.4	185437	2	AC012075	Human sapi	2	AC012075	Human sapi		
15	39	12.4	218939	2	AC079997	Human sapi	2	AC079997	Human sapi		
16	38.6	12.3	110000	2	PFMAL13P2_2	Continuation (3 of	2	PFMAL13P2_1	Continuation (3 of		
17	38.4	12.2	195761	2	AC018895	Human sapi	2	AC018895	Human sapi		
18	38.4	12.2	203257	2	AC013737	Human sapi	2	AC013737	Human sapi		
19	38.2	12.1	124281	2	AL136101	Human DNA	2	AL136101	Human DNA		
20	38.2	12.1	173452	2	AC003013	Human PAC	2	AC003013	Human PAC		
21	38	12.1	9577	1	U67580	Methanococ	1	U67580	Methanococ		
22	38	12.1	68464	2	AC005248	Human sapi	2	AC005248	Human sapi		
23	38	12.1	169469	2	AC037195	Human sapi	2	AC037195	Human sapi		
24	38	12.1	172144	4	AC092497	Sus scrofa	4	AC092497	Sus scrofa		
25	38	12.1	175946	2	AP002088	Human sapi	2	AP002088	Human sapi		
26	38	12.1	181686	2	AC097819	Rattus no	2	AC097819	Rattus no		
27	37.8	12.0	120984	2	AC006463	Human sapi	2	AC006463	Human sapi		
28	37.8	12.0	175735	2	AC068566	Human sapi	2	AC068566	Human sapi		
29	37.8	12.0	192120	2	AC104640	Human sapi	2	AC104640	Human sapi		
30	37.6	11.9	14059	3	AE001390	Plasmodiu	3	AE001390	Plasmodiu		
31	37.6	11.9	36255	2	AC016208	Human sapi	2	AC016208	Human sapi		
32	37.6	11.9	88037	2	PFMAL13P8	Continuation (3 of	2	PFMAL13P8	Continuation (3 of		
33	37.6	11.9	146389	2	AC023916	Human sapi	2	AC023916	Human sapi		
34	37.6	11.9	166703	2	AL139378	Human sapi	2	AL139378	Human sapi		
35	37.6	11.9	167257	2	AC008808	Human sapi	2	AC008808	Human sapi		
36	37.6	11.9	167553	2	AC090042	Human sapi	2	AC090042	Human sapi		
37	37.6	11.9	169480	2	AC068074	Human sapi	2	AC068074	Human sapi		
38	37.6	11.9	182651	2	AC104580	Human sapi	2	AC104580	Human sapi		
39	37.4	11.9	104089	2	AC012488	Human sapi	2	AC012488	Human sapi		
40	37.4	11.9	112212	2	AL136096	Human sapi	2	AL136096	Human sapi		
41	37.4	11.9	139226	2	AC084233	Human sapi	2	AC084233	Human sapi		
42	37.4	11.9	166946	2	AP001008	Human sapi	2	AP001008	Human sapi		
43	37.4	11.9	222029	2	AL606831	Mus muscu	2	AL606831	Mus muscu		
44	37.2	11.8	12240	2	AE001424	Plasmodiu	2	AE001424	Plasmodiu		
45	37.2	11.8	167491	2	AC093442	Papio cyn	2	AC093442	Papio cyn		

## ALIGNMENTS

AL356798 118068 bp DNA linear PRI 26-SEP-2001  
Human DNA sequence from clone Rpl1-467B11 on chromosome 9, complete sequence.  
AL356798  
AL356798.18 GI:15795413  
HTG.  
ORGANISM  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE  
1 (bases 1 to 118068)  
Babbage,A.  
Direct Submission  
Submitted (26-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
requests: clonerequests@sanger.ac.uk  
On Sep 27, 2001 this sequence version replaced gi:14455883.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep). This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/NGP/Ch9>. RP11-467B11 is from the library RPCT-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/Bacpac/home.htm>. VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-467B11. It may be shorter because we sequence overlapping sections only once, except for a short overlap. The true right end of clone RP11-467B11 is at 118068 in this sequence. The true left end of clone RP11-208F1 is at 29578 in this sequence. The true right end of clone RP11-29B11 is at 2000 in this sequence.

FEATURES  
source  
1. 118068  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="9"  
/clone="RP11-467B11"  
/clone\_lib="RPCT-11.2"  
BASE COUNT 33053 a 23625 c 22510 g 38880 t  
ORIGIN

Query Match 100.0%; Score 315; DB 9; Length 118068;  
Best Local Similarity 100.0%; Pred. No. 1.3e-62;  
Matches 315; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taacacgactagaatttaagaataaacttttggaataatagaacaatgaatgaatt 60  
|||||  
Db 96845 TAACACTGACTGATGATTTAAAGAAATTAAGTTCGAGAAATGACAAATGAAATCAGTT 96786  
QY 61 tctccaccacttaagatatctcttagagatctacagcctcccttagggagacatacaa 120  
|||||  
Db 96785 TCTCCACACTTAAGATATATCTTTAGAGATCTACAGCCTCCCTTTAGGGGACATACAAA 96726  
QY 121 gtcagttgtgtgcttctttagtcacacactatataatgaatgagtagatgactacaatt 180  
|||||  
Db 96725 GTCAGTTGTTGCTTGTGTTGAGTCCACCTTATTTCAAGTAGTATGACTACCAATT 96666  
QY 181 ttgaataatgattgtacacataaactcgtgagtttatggaaacatcagtagaagaataa 240  
|||||  
Db 96665 TTGAATAATGATTGTACACAAATAAAGTTCGAGTATGGAACATCAGTAGAAGGAATA 96606  
QY 241 caacattccatcccttaagaagatcattacttgaactcagatgaatatttcattgtgt 300  
|||||  
Db 96605 CAACATTCATCCCTTTACAGAGATCATTTACTTGCAACTCAGGATTAATTTGTCAATGT 96546  
QY 301 attatctactatgc 315  
|||||  
Db 96545 ATTATCTACTATGC 96531

RESULT 2  
AC025794 AC025794 226399 bp DNA linear HTG 13-JUN-2001  
LOCUS  
DEFINITION Mus musculus clone RP23-6207, WORKING DNAFT SEQUENCE, 25 unordered  
1

ACCESSION  
AC025794  
VERSION  
AC025794.4 GI:14389364  
KEYWORDS  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE  
house mouse.  
MUS MUSCULUS  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 226399)  
Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
Mus musculus, clone RP23-6207  
Unpublished  
2 (bases 1 to 226399)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choquel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Laroque, K., Lamazares, R., Landers, T., Lehotzky, J., Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., Margulis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pisanil, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talama, J., Testaferr, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zairoun, J., Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
COMMENT  
Submitted (14-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jun 13, 2001 this sequence version replaced gi:8225449.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: sequence.submissions@genome.wi.mit.edu  
----- Project Information -----  
Center project name: L8057  
Center clone name: 62\_0\_7  
----- Summary Statistics -----  
Sequencing vector: M13; M7815; 45% of reads  
Sequencing vector: Plasmid; n/a; 55% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 217695 bases at least Q40  
Consensus quality: 221726 bases at least Q30  
Consensus quality: 223237 bases at least Q20  
Insert size: 219000; agarose-fp  
Insert size: 223999; sum-of-contigs  
Quality coverage: 8.2 in Q20 bases; agarose-fp  
Quality coverage: 8.0 in Q20 bas.  
NOTE: This is a 'working draft' sequence. It currently consists of 25 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  
1 15086: contig of 15086 bp in length  
\* 15087 15186: gap of 100 bp  
\* 15187 15923: contig of 737 bp in length  
\* 15924 16023: gap of 100 bp



```
*
* 16024 16694: contig of 671 bp in length
* 16695 16794: gap of 100 bp
* 16795 17880: contig of 1086 bp in length
* 17881 17980: gap of 100 bp
* 17981 18573: contig of 589 bp in length
* 18580 18679: gap of 100 bp
* 18680 19268: contig of 589 bp in length
* 19269 19368: gap of 100 bp
* 19369 20014: contig of 646 bp in length
* 20015 20114: gap of 100 bp
* 20115 20821: contig of 707 bp in length
* 20822 20921: gap of 100 bp
* 20922 21533: contig of 612 bp in length
* 21534 21633: gap of 100 bp
* 21634 22538: contig of 905 bp in length
* 22539 22638: gap of 100 bp
* 22639 22338: contig of 700 bp in length
* 23339 23438: gap of 100 bp
* 23439 24125: contig of 687 bp in length
* 24126 24225: gap of 100 bp
* 24226 24941: contig of 716 bp in length
* 24942 25041: gap of 100 bp
* 25042 25287: contig of 846 bp in length
* 25288 25987: gap of 100 bp
* 25988 26876: contig of 889 bp in length
* 26877 26976: gap of 100 bp
* 26977 28437: contig of 1461 bp in length
* 28438 28537: gap of 100 bp
* 28538 30287: contig of 1750 bp in length
* 30288 30387: gap of 100 bp
* 30388 32661: contig of 2274 bp in length
* 32662 32761: gap of 100 bp
* 32762 36782: contig of 4021 bp in length
* 36783 36882: gap of 100 bp
* 36883 45293: contig of 8411 bp in length
* 45294 45393: gap of 100 bp
* 45394 61819: contig of 16426 bp in length
* 61820 61919: gap of 100 bp
* 61920 126287: contig of 64368 bp in length
* 126288 126387: gap of 100 bp
* 126388 149055: contig of 22668 bp in length
* 149056 149155: gap of 100 bp
* 149156 211213: contig of 62058 bp in length
* 211214 211313: gap of 100 bp
* 211314 226399: contig of 15086 bp in length.
*
FEATURES
    source
        1..226399
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone="RP23-6207"
            /clone_lib="RPC1-23 Female Mouse BAC"
            1..15086
                /note="assembly_fragment"
                clone_end:SP6
                vector_side:left"
            misc_feature
                15187..15923
                    /note="assembly_fragment"
            misc_feature
                16024..16694
                    /note="assembly_fragment"
            misc_feature
                16795..17880
                    /note="assembly_fragment"
            misc_feature
                17981..18579
                    /note="assembly_fragment"
            misc_feature
                18680..19268
                    /note="assembly_fragment"
            misc_feature
                19369..20014
                    /note="assembly_fragment"
            misc_feature
                20115..20821
                    /note="assembly_fragment"
            misc_feature
                20922..21533
                    /note="assembly_fragment"
            misc_feature
                21634..22538
                    /note="assembly_fragment"
            misc_feature
                22639..23338
                    /note="assembly_fragment"
```

```
misc_feature 22639..23338
              /note="assembly_fragment"
misc_feature 23439..24125
              /note="assembly_fragment"
misc_feature 24226..24941
              /note="assembly_fragment"
misc_feature 25042..25887
              /note="assembly_fragment"
misc_feature 25988..26876
              /note="assembly_fragment"
misc_feature 26977..28437
              /note="assembly_fragment"
misc_feature 28538..30287
              /note="assembly_fragment"
misc_feature 30388..32661
              /note="assembly_fragment"
misc_feature 32762..36782
              /note="assembly_fragment"
misc_feature 36883..45293
              /note="assembly_fragment"
misc_feature 45394..61819
              /note="assembly_fragment"
misc_feature 61920..126287
              /note="assembly_fragment"
misc_feature 126388..149055
              /note="assembly_fragment"
misc_feature 149156..211213
              /note="assembly_fragment"
misc_feature 211314..226399
              /note="assembly_fragment"
              clone_end:77
              vector_side:right"

BASE COUNT  61164 a 51232 c 50642 g 60960 t 2401 others
ORIGIN
```

```
Query Match      12.8%; Score 40.4; DB 2; Length 226399;
Best Local Similarity 54.8%; Pred. No. 5.6;
Matches 80; Conservative 0; Mismatches 66; Indels 0; Gaps 0;
```

```
OY 10 actcagatttaagaacttcttgaagaataagaacaatgaatcagttccaccac 69
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 51932 ATTCGCCATTTCGGAAATGCAATTCACACAGTCTTCAGATGAAATGCTTTCACACTA 51991
OY 70 ctttagtatctctcttagagatctacagctcccttccttgaggacataaagtcagttgt 129
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 51992 CTGAACAATACCTCAAAATATTCACCTTACCACACCACTTGATGACCAATTCAGTGGT 52051
OY 130 gtgccttggtgagtcaccattat 155
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 52052 CTTGGCCTTGAAANAAGCCAGTGTCAAT 52077
```

```
RESULT 3
AF429315 125020 bp DNA linear PRI 18-JAN-2002
LOCUS DEFINITION Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION AF429315
VERSION AF429315.1 GI:17646244
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 125020)
            Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
            Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
            Potter,N.T., Ross,C.A. and Margolis,R.L.
            A repeat expansion in the gene encoding junctophilin-3 is
            associated with Huntington disease-like 2
            JOURNAL Nat. Genet. 29 (4), 377-378 (2001)
            MEDLINE 21583737
            PUBMED 11694876
```



Oy 93 tacagcctcccttaggggacatacaagtcagtgctgtgcttctgttgagccacct 152  
 Db 67292 TTCACCTTCACTAAGGATACATCATCTTGACTCTACTCATAGATCATCTTGCT 67351  
 Oy 153 tatatcaagtagtgatgactacaatttgaaatagatgct 195  
 Db 67352 GTTATTGAACCTTATATCATATCATATATATATATATCTCT 67394  
  
 RESULT 5  
 AC102444/c 62871 bp DNA linear HTG 23-NOV-2001  
 LOCUS Mus musculus clone RP24-196115, LOW-PASS SEQUENCE SAMPLING.  
 DEFINITION AC102444  
 ACCESSION AC102444.1 GI:17061530  
 VERSION HTG: HTGS-PHASE0.  
 KEYWORDS house mouse.  
 SOURCE Mus musculus  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 62871)  
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
 JOURNAL Mus musculus, clone RP24-196115  
 REFERENCE 2 (bases 1 to 62871)  
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
 Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Bouckgalter, B.,  
 Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,  
 Chepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,  
 Cooke, P., Dearrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fato, S.,  
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,  
 Girdle, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
 Haas, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
 Jones, C., Kanat, A., Karatas, A., Kells, C., Laroque, K.,  
 Lamazares, R., Landers, T., Lenockzy, D., Levine, R., Liu, G.,  
 Maclean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C.,  
 McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J.,  
 Menes, L., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C.,  
 Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
 Oliver, J., Peterson, K., Phunhkanh, P., Pierre, N., Pollara, V.,  
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,  
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
 Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,  
 Vialon, J., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, M.J., Young, G.,  
 Zainoun, J., Zemdek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: MIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L18811  
 Center clone name: 196\_1.15  
 -----  
 \* NOTE: This record contains 84 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

1 626: contig of 626 bp in length  
 627 726: gap of 100 bp  
 727 1387: contig of 661 bp in length  
 1388 1487: gap of 100 bp  
 1488 2127: contig of 640 bp in length  
 2128 2227: gap of 100 bp  
 2228 2865: contig of 638 bp in length  
 2866 2965: gap of 100 bp  
 2966 3591: contig of 626 bp in length  
 3592 3691: gap of 100 bp  
 3692 4331: contig of 640 bp in length  
 4332 4431: gap of 100 bp  
 4432 5087: contig of 656 bp in length  
 5088 5187: gap of 100 bp  
 5188 5836: contig of 649 bp in length  
 5837 5936: gap of 100 bp  
 5937 6580: contig of 644 bp in length  
 6581 6680: gap of 100 bp  
 6681 7335: contig of 655 bp in length  
 7336 7435: gap of 100 bp  
 7436 8093: contig of 658 bp in length  
 8094 8193: gap of 100 bp  
 8194 8836: contig of 643 bp in length  
 8837 8936: gap of 100 bp  
 8937 9589: contig of 653 bp in length  
 9590 9689: gap of 100 bp  
 9690 10312: contig of 623 bp in length  
 10313 10412: gap of 100 bp  
 10413 11061: contig of 649 bp in length  
 11062 11161: gap of 100 bp  
 11162 11813: contig of 652 bp in length  
 11814 11913: gap of 100 bp  
 11914 12561: contig of 648 bp in length  
 12562 12661: gap of 100 bp  
 12662 13507: contig of 646 bp in length  
 13508 13407: gap of 100 bp  
 13408 14046: contig of 639 bp in length  
 14047 14146: gap of 100 bp  
 14147 14802: contig of 656 bp in length  
 14803 14902: gap of 100 bp  
 14903 15545: contig of 643 bp in length  
 15546 15645: gap of 100 bp  
 15646 16288: contig of 643 bp in length  
 16289 16388: gap of 100 bp  
 16389 17032: contig of 644 bp in length  
 17033 17132: gap of 100 bp  
 17133 17798: contig of 666 bp in length  
 17799 17898: gap of 100 bp  
 17899 18526: contig of 628 bp in length  
 18527 18626: gap of 100 bp  
 18627 19296: contig of 670 bp in length  
 19297 19396: gap of 100 bp  
 19397 20033: contig of 637 bp in length  
 20034 20133: gap of 100 bp  
 20134 20764: contig of 631 bp in length  
 20765 20864: gap of 100 bp  
 20865 21496: contig of 632 bp in length  
 21497 21596: gap of 100 bp  
 21597 22252: contig of 656 bp in length  
 22253 22352: gap of 100 bp  
 22353 22994: contig of 642 bp in length  
 22995 23094: gap of 100 bp  
 23095 23763: contig of 669 bp in length  
 23764 23863: gap of 100 bp  
 23864 24520: contig of 657 bp in length  
 24521 24620: gap of 100 bp  
 24621 25282: contig of 662 bp in length  
 25283 25382: gap of 100 bp  
 25383 26040: contig of 658 bp in length  
 26041 26140: gap of 100 bp  
 26141 26795: contig of 655 bp in length  
 26796 26895: gap of 100 bp  
 26896 27535: contig of 640 bp in length

Query Match	12.6%	Score 39.6;	DB 2;	Length 62871;
Best Local Similarity	49.5%;	Pred. NO. 11;		
Matches 102; Conservative	0;	Mismatches 104;	Indels 0;	Gaps 0;

RESULT	6				
AF318573		AF318573	108873 bp	DNA	linear
LOCUS		Bovine herpesvirus 4	long unique	region,	complete sequence.
DEFINITION		AF318573 AF271211			
ACCESSION		AF318573.1	GI:12802528		
VERSION					
KEYWORDS					
SOURCE		Bovine herpesvirus 4.			

REFERENCE  
AUTHORS  
Zimmermann, W., Broll, H., Ehlers, B., Buhk, H.-J., Rosenthal, A. and  
1 (bases 1 to 108873)  
Schmiedel, P. and Lind, C.

TITLE	Genome Sequence of Bovine Herpesvirus 4, a Bovine Rhadinovirus, and Identification of an Origin of DNA Replication
JOURNAL	J. Virol. 75 (3), 1186-1194 (2001)
PMID	11152491
REFERENCE	2 (bases 1 to 108873)
AUTHORS	Zimmermann, W., Broll, H., Ehlers, B., Buhk, H.-J., Rosenthal, A. and Gales, M.

**JOURNAL** Direct Submission  
Submitted (03-NOV-2000) Project P24, Robert Koch-Institut, Nordufer  
20, Berlin D-10555, Germany  
**FEATURES** Location/Qualifiers  
1..108873  
source

```
misc_feature 1. .108873
CDS          /note="long unique region; LUR"
             complement(445. .633)
```

```

/evidence=not_experimental
/product="hypothetical protein pb01"
/protein_id="AAK07921.1"
/db_xref="GI:12802529"
/translation="MEGVDSVAAGRRRTRODLNNSAGNEVCCAVLRKHNYSSAGC
CAVRNNSFHHSPLGMP"
614. .4480
DSDS

```

```

Query Match      12.6%; Score 39.6; DB 2; Length 62871;
Best Local Similarity 49.5%; Pred. No. 11;
Matches 102; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

      *      50082 50181 gap of 100 bp in length
      *      50182 50843: contig of 662 bp in length
      *      50844 50943: gap of 100 bp
      *      50944 51589: contig of 646 bp in length
      *      51590 51689: gap of 100 bp

/translation="MDTSPSIDLCEFFVRRAPLNEECVIOQFSDCOLPTPTSVYLV
EQVFVNLVNGEEOAVRFLVPLDRLKEFSGHGLFFHRSRLIKTSDRSLPTGYGSPA
PEFALSRLPVYALGMSGLAPMSNLTOLIKVLSITISICNINPARKHMTLRL
CPPSLDFRPYKSELYRHANHTHAPQSVLSPRPGLDLGAEVGFVLNAGSAGSEFMT
PHQGVQTIYNQFCKTATIVPEVDWIPETISYPLSSGHMMDTGLATYVGGLOVAPRN
LIKLFYNNQMFQNGFVNFYSQGLASIVSIITPSSRSSELTLPYRTRMDVYMNMRCC
SLPCTQGYFFPVLVESHLPLDSVPEGLIATGAVCFSRNITLDPLEPRHKNINYOGLT
FTSSHEDSGRGRECLGALKYIMSELRPOYKVGISAVKEVEMAHKALISVCGDTYGIK
IYOSALPPLNLTGLPPTAASRNQMMIMIKOHPINIAVAPVATAPEPTDPTDYLYK
ACGALHGCQVYVIGRRQCKPHIITLKHQKEDQENVERKISLKRSPSYFLSTNNLQYK
ENMOSSKHEMPTLBSLTLEILIKHEPVGCKDITVHIDILDSGVRQ00QIGIADII
PISDSIIVSDISLAAABDRQSENPMVAPKEDLDLPLVNNPEPAGCSAIGCEPLSI
TYEPIKRGAAIVESLNLIVAPKRIIDVYQCFENITVAPQDCTTSHFGIILEMAGCEE
CSQIIGCVFVPSCTSNRQCKLNNSIVRIVITVATAPQCDVYQGTPLKRPEDDSI

```

CDS	WLPSTEVHTFETVTSOLFRDPSPSGHIINIDPLVYKKLTIVTKHLINHESIISCHDV GGGLITAEKEMAYAGASIALTVPODDPVFLFSETPGGLAEVPRKAVSTVOXKLE TSDIIFDVAGRTLPVSANFTETVSYKREBITPRESINEMVENNRHSTYKQRTYCEY OGTPEDPKDQJLHITFOPSMCPHGPYTHOVNVLTPGINPDSLVALEBAGFRVN LVSTFETDKVATIDVTIVFGLIGANINIBATLGDKAISMYTKHNSVLGELKLI NSPDEFSLAIGHTAQOILFENKFMGNKPSNTYCKENYSGIMESRWLNFEIPENTH VALQSMKSGSLPGWITQGHGFAHSELYEMMLSTHGMVATQFAGDISAGPALTYP ONPTAGTISGLCSADGRHLALHDPGSLNNLMQMPHLPKMPPLAKVSPWKRMFLDLH IWANKVREMDQPPPHDPDLNRKIKYM"
CDS	4608..4811 /note="ORF B02; B0RFA2" /codon_start=1 /evidence="not_experimental" /product="hypothetical protein pB02" /protein_id="AAK07923.1" /db_xref="GI:12802531"
CDS	/translation="MASPPQYCHRMVRYHGPGRKSAKISYGVKKDRDYTHQLSKNSAYS -GMSPPQSCADLNKLFVSQCPIS" 5286..5513 /note="ORF B03" /codon_start=1 /evidence="not_experimental" /product="hypothetical protein pB03" /protein_id="AAK07924.1" /db_xref="GI:12802532"
CDS	/translation="MGFNPVMLYHPACNCTPISLKHGILYIKIKISYLLIRATFLRVN HIGPKRDRTPGCGRPPSLIVQRLALK" 5943..9347 /note="ORF 6" /codon_start=1 /product="single-stranded DNA-binding protein MDBP" /protein_id="AAK07925.1" /db_xref="GI:12802533"
CDS	/translation="MAAKTAQOQOYEDINGSKAPRGPGCIYIVYKKEFPFKREASLIG NFPGETSVASVPLFGLVEPGLINVAHKKIDTTVYSAKVITSYHREVMESNIDC FTPIFHGQPLQCOETRDLEGFTHSPSPSNTNNISLQDLGTCNCDIEWLIGVIT EGFKERLYHGHIPLGLHVEEVOISPOHAAKRIPLYDEDLFSKIAPDMKRFHSYVS OYLFNSLYTAIQAQIRVKDVATVIOAMESOFRODHMKPKVYHKRDPSSRSRPGDL SLMIDVASLEAVSYGLSFIDAPQDMASALIDYTSMPITFVOCITEEDRLQAHAMNK OSIHYNQOLSTNSMLYTRIDOKONPKTRAKGDLANYNSYILOHGISYSEARQDXYGQ PVFGVOGNLSGTYTTHHLATVPAASMCENTLARYCYILOFCOHKSTONOSYNSNY VGTAAASDMCNLCQGRKPAVCINTLFLYRLRDRFPPIITSHRDPYITIGSVAMYNLD ILGNFASFREKEDEGAQVEVQKTYWQTLTLEKLEGMGKIDTPPEKLTSLDIPSF IKVEFDIDAHVDEVLKFINCKAVNNINERETIKSIHHYIOYCCNVFAOPPCPVLOL YKKSITLIIDYDCLPNCMMYEDDNSSMGPEMLKMYOTLMTNKKGCCPRGVLIG AEIKYMSHDPPDFPDVDAINGISPTKTOVRISRAMIYVPTTIKIKRRIIFSSNSG SEAIOSGLFKSGSKDNYITVETPKMFLNITHKIMFPQTKSSLFMMHFTSTKQIPL ITGVPRQDLVALANIEYNSKLHSEIDVLDIIPDMLTYAKIKLNNAILRACGQOQY STTHCLVPTVATVGEERYPHVLENEEFGTEQYLAQONRTAQIVOATLKEDVAMQ KLRLIITVPMVINKYTGINGNNGIFHCGNLGYFMRGVDRLIFENAPPKRSTNAYM RKHYFMTPIYDNLIKRTVPPSPSTFEVENTIRDRDMLLEDKONONIFPDVYLELVKG LGOACADIFADIDQFYIGEYVYMSDEILSRQISIDAGVPMCEDSVSSFLGELYTSEE OLEFTGLSEOTTSAPTSEDFQASGLSTIAAGKRRKLKMLCMSDFDL"
CDS	9352..11430 /note="ORF 7" /codon_start=1 /product="transport protein" /protein_id="AAK07926.1" /db_xref="GI:12802534"
CDS	/translation="MARELAAYSQVDFLSFKTSLITFCDPRHIDLGAFQRRKRLKL LIDALYSHLVEONRTESCISLELHLIINTYITMQKLETENINIEYDPAAMLHS DLCEHKTIVNLCFYGDCVPISTIINDIEFRLNSVFCISNNALLEGLEGTINF LGLLRNISPRLPDLYPSISCMNLNENSLIPNGENAVFTMNSINCHVATPYAPE PYQGLFENELMHLNTPVDGTDKVOLETSDRITQNASIEAIDOHNIIEKVSCHILSIS NLITYNSGSHNDQSPRLSCSMATLIQHTTRKASRQVASHALATTPRIHFEDSYRA YPIESLFCGSIYSIMDNIQALQDCSTFLFASNYKTIEMRKNELYVANNLLASH TKESGIDKPNISIKOARHTTIGETTSEOIFSAARKNNYLOKVTTEGKRLTDLDET OGKIMCNTLLTRTGMVYVKEAAILKNEHMIRKRIALDPMHNTYTAINEYESKFI KNSLIYVOKISOEHIDSTIQYSLITGPIVESPDPFPSPRNILLANCDAAVYMHKK MLISEVIMPSMEKDMIDAEFNAPYNTGKTLNSVQOQAMRYIRELVMSVSLYNNWE KDLNIVSPSTLSPHNLFSBKPFKSGIYIYTBKAPLFIHNTGMIFKDLVALLYHH IOLTSHKNVL"
gene	11420..14044
CDS	/gene="gb" 11420..14044 /gene="gb" /note="ORF 8" /codon_start=1 /product="glycoprotein B" /protein_id="AAK07927.1" /db_xref="GI:12802535"
CDS	/translation="MYKTLIFALIKVCSFNQTTHTSTTSPSISSTTSSTSTSK PSNTYSTNSLSLAOSPNTSKSPSTDNQGTSTPIPTVYTDASNNKTRVCSASS SGELRFDLQDPCPDPKKHVGEILLVLAKNIIVYIKRVKRYRRIATISVYRWGMSO AAVTNRDDISRAIPYNEISMIDRIYHCSAMATYINGLINTYIDRSENKSVLPQPA GLTENINIRFESQPLIYABEGMPGPIYRRTVNCVVMVMAVSAPYHPIFALDPTI EISPRCHNNSOCTGTGNSRQATKYMIENHQTVDYERRGHPTKDKRIPIKDEEYTI WKAEDREAIQDFYIMKTFPPRAIQTIINHSEFPVANEVYASLTSNOEETELRGTEI LNCMNSTINETLEETVAKFMNSHIDGEEKYIKINGLFLINQAKPPLASLHNTYTI ERNNTGNKSRQKRSVDKRTQDAGSLSTAQOYAVDHLRTSMNHILBELTPTWCREQ KDNLMWYELSKINPVSMAALYGRVAVKAGDAFVNSSECINVOASVNIHKSMTD DPKVCYSRPLVTEFVNSTAFPRGQIGTRNELINTHTVETCRPTADHYFVKNMTHY FKDYKFEVMTQTNNTSLDPTFLTLNLFEDINDFTVELYSTEEKMASALDETFMR EYNYTOKLASREDNDNTIDNRDLKQDISEMAADGQIDGKVVYVNFSGIYVYFS IVGCVSFEFTNIDIGVTTILLITIVYVVFYTSKPTNNKNEAPIYMTIPNDKASEOE NIQPLGGEIKRIILGMHQLOOSEHGKSEESKSPGLFOLLGDELQILRRGYRLP TFDPSGNDTSETHQYV"
CDS	14184..17201 /gene="Dpol" 14184..17201 /gene="Dpol" /note="ORF 9" /codon_start=1 /product="DNA polymerase" /protein_id="AAK07928.1" /db_xref="GI:12802536"
gene	/translation="MEFNPIYSKKTDRKISCVDDPDKTNPYRLVLPCKPTPTGP GVISTITTDTPVLFHQDKVEPIETSGDHQORSWSALKPPAIPKDKLFFVYDILIE TIYTPDRPDPIEQOTDIPNGVTLKILGQDQDASVCNVFQOMVYVYLVPGVN LSYVLOQTLGVNGVNTQCKFSITQORRKILKRYDLSLIPVYITISAPTEINQALGNL TSCGSEVESVNSASTRFIDNKFSTFGVSGSNPHRSTQSDSTFDEPGLDLO FHEEEOOMPRTYIMSDPIEDCIGEGFPQATKREDIVIOISCIIMVYSCSSPNNILS LQTCPIANTDYERPESELDNFATFTLLRDNIDFVYGINYANRDPITIDRATQY NIPRHFRTLVISGSTEFEVQNPMTGAGFERSKSIKISGLVYIDMTQYCKDLMSDY KLNTYAKHQMGOEKEDVYKELPSLRSGEAGARIGSYCVLDSVLYDLKYFM1HV EISELAKMAKIPARVLYLDGQOIRYFSCILLEAKKEHFLIPRGAQKPGQXGATVIN PRPGYVPIIIVDPAASYPSIIQAHNCLYSPIODONLHLHLKRPDYETFHLSTGP IHEVQKHCTSLSTLLTAMLAUKRKRTIKELACDQGPMTIILDRQOLAIKYTCMSVY GFTYASGILCPICPIAETVYTDQRTMLESKAFVEMITPERISDIYVSPVDPDPAF RVITYDTSLETECGYPMQSVLNFADKLFTYTAALFEDPIIKLAERKFOCLMLAK KRYIGILSSDKLVMGVLDLRNTAASFEVQTSKEILDVLRDPEVQAQOAYLCRQAPA KYVSGLPEVGEKVIDVLNNSYAAALATGVSVESLTFSTELSPRGEYVTTLPILHTV YKIRSNREBELFOIDRIIPYVFIKQDORCKEDLAEDPPYVQNKIPLISVEYFDKLI HGVANILQCLFETGNSNMVYELILYNNIYSPST"
CDS	17297..18577 /note="ORF 10; B0RFB1; conserved in other gamma-herpesviruses" /codon_start=1 /product="unknown" /protein_id="AAK07929.1" /db_xref="GI:12802537"
CDS	/translation="MTSLVERFSISGMOVNIONGNIIIMTNLKDIPVQGTGMLYKVV FLPLSLQDINNYISRGVQASIFELEDAQACYAMQVEFPPIRTPLDETPIHVEFN MPLFLVQONTIINSLSLAIIPMKPTCPGASISYSTAPEPIPOHSHNLHPLS TDQGHLLTSGAIIKGNHKTYTVPFRGSPMOEFGMHTVASGRDLODACVGMVSYD IKNIMAEVYNPAGIKYKVSNTINPKRPHQIWTTLKFTYVWAKTQDLIVASVPSSTLPS CGKVNNTYDSEKILKPGSTLLMLKIKYTTTRGNESSTKAVMPTTGTNTNPMVITIEPTIM LPMTPLOVTKIKPMIIMTITIKKDLAIAACVPRYSSTLEDRQPPASVYFNPQDLITTW EDSMVNTAGENIIYSRCHLNINSENTSPMDPT"
gene	complement(18848..19345) /note="ORF B04; short ORF of immediate early transcript 1 RNA" /codon_start=1
Query Match	12.6%; Score 39.6; DB 14; Length 108873;
Best Local Similarity	50.4%; Pred. No. 9.6;

Matches 125; Conservative 0; Mismatches 119; Indels 4; Gaps 1;				FEATURES	
QY	32	tttgagaataagaacaatgaatcagttcttcaccacttatactcttagagat	91	source	overlap between neighbouring submissions.
Db	49499	TTTGAGATGTGAGCGATATAGTCAAAAGTATATTAAATCAATATATTATCCACACGCC	49558		Location/Qualifiers
QY	92	ctacaagcctcccttgaggagacatacaagtcagttgtgttccttcttgatccacc	151		1..54700
Db	49559	TAACTGTATTGATTATAGCATTCATGCAAAAGTCCCTCAGATGGAATTTTGGAGTCTCT	49618	gene	/organism="Caenorhabditis elegans"
QY	152	tatatatt---caagtaggtaagtaactaatcttgaataatagatgtgcacaataaac	207		/db_xref="taxon:6239"
Db	49619	TGATTTTGTACTAAATATCTCATGATGAAATGGAATGTGATTTTGAACCGACTG	49678	CDS	/chromosome="IY"
QY	208	tggagtttgaagaacatcagaaggaataacacatccatcccttcagaagatca	267		/clone="Y116A8A"
Db	49679	TTCGTGTTTGAATTAATTAATGATTTTAAGTATATGTTTCCAAATCTGATGATCC	49738	gene	/gene="Y116A8A.1"
QY	268	ttactattg 275			join(1517..1562,1674..1768,3071..3330,3372..3516)
Db	49739	TTTGATG 49746		CDS	/gene="Y116A8A.1"
RESULT 7					join(1517..1562,1674..1768,3071..3330,3372..3516)
LOCUS	CEY116A8A	54700 bp	DNA	linear	/note="predicted using GeneFinder"
DEFINITION	Caenorhabditis elegans cosmid Y116A8A, complete sequence.				/codon_start=1
ACCESSION	AL117205.2	GI:6425237			/protein_id="CAB55161.1"
VERSION	HTG.				/db_xref="GI:5832824"
KEYWORDS	Caenorhabditis elegans.				/translation="MLHRTNASCILGHEFDYGSYELSTFPSINILKSCQIIYFAVNP
SOURCE	Caenorhabditis elegans.				VNPVYASKLITLAQACTINAGSLGSLQNDKERWVIGNKGMETIMLANGTDSGTWIG
ORGANISM	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.				AKCSSTGCTWTDGNTVGTQGMFAFPGENPDLSTPPCLYIAWIKIGDTLKPNGNYID
REFERENCE	1 (sites)				DIRCTTAMMSYACGKPGLRNN"
AUTHORS	none.				complement(join(4463..4607,4651..4910,5667..5876))
TITLE	Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium				/gene="Y116A8A.3"
JOURNLM	Science 282 (5396), 2012-2018 (1998)				/codon_start=1
REMARK	99069613				/protein_id="CAB55165.1"
REFERENCE	2 (bases 1 to 54700)				/db_xref="GI:5832828"
AUTHORS	McMurray/A.A.				/db_xref="SPRREMBL:O9U2M4"
JOURNAL	Direct Submission				/translation="MKRLIOLIAFCFLIIPASAFIRDPFSSSSEEGYGRCKHHNG
COMMENT	Institute (31-DEC-1998) Nematode Sequencing Project, Sanger Institute, Hinxton, Cambridge CB10 1SA, England and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jles@sanger.ac.uk or tw@nematode.wustl.edu				HHSPTNPCEWGYAFHSPGQVWCVKVAASKLIVLAQACTINAGSLQNDKERY
	On Nov 15, 1999 this sequence version replaced gi:5832821.				WIGNKGMETIMLANGTDSGTWIGAKCSSTGCTWTDGNTVGTQGMFAFPGENPDLSTPPCLYIAWIKIGDTLKPNGNYID
	Coding sequences below are predicted from computer analysis, using Predictions from GeneFinder (P. Green, U. Washington), and other available information.				CLYIAWIKIGDTLKPNGNYID
	Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones.				complement(join(12804..13270,13318..13360))
	Exceptions are indicated by an explicit note.				/gene="Y116A8A.2"
	IMPORTANT: This sequence is not the entire insert of clone Y116A8A. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.				complement(join(12804..13270,13318..13360))
	The true left end of clone Y116A8 is at 1 in this sequence. The true left end of clone T06A10 is at 54597 in this sequence. The start of this sequence (1..106) overlaps with the end of sequence AL13846.				/gene="Y116A8A.2"
	The end of this sequence (54597..54700) overlaps with the start of sequence AT099926.				complement(join(4463..4607,4651..4910,5667..5876))
	For a graphical representation of this sequence and its analysis see: <a href="http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=Y116A8A">http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=Y116A8A</a>				/gene="Y116A8A.3"
	IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small				/codon_start=1
					/protein_id="CAB55161.1"
					/db_xref="GI:5832825"
					/translation="MKFQWITGLVLTIGVINCILSSQKLTIVENERLKSQNTLPIP
					SINQTSKGGCFPNDRKNTITSNSGWLILAGHRNDRSTRYATAAFPISEYHILISRWML
					TEHKWIKNGKRPDKNNCSGGRHLDLPVPRVDNLTLEWYGRKPKRADSKAPFACVNC
					ODKAPVILIELNKTSKRWANWGPVCLADVEDTSKLDVWVFYGSYDYLHKLLI
					KAVKDGGITETIYETIEVGGPLVONWVSKAATYIGAKAGSRNCELFYDLSKLEKIC
					DYSGVGYEKPTLTTEPPTETTERKSTVSTTISTSELTJIMASSASPSIKPSTTTP
					SPLQASSKRAPEPLIRPSRERRASSASRPVQTTITPELTISLPPPELKTSTVT
					TTTTPEATYDVDDDMIGLDETDENSKVGVKIDRVGFVLLVIE"
					complement(join(25359..25414,25792..26337,26389..26506,
					26554..26633,26676..26793,26862..26943,26989..27235,
					27279..27402))
					/gene="Y116A8A.6"

CDS  
 complement(join(25359..25414,25792..26337,26389..26506,26534..26633,26676..26793,26862..26943,26989..27235,27279..27402))  
 /gene="Y116A8A.6"  
 /note="predicted using GeneFinder  
 cDNA EST yk82d6.5 comes from this gene  
 cDNA EST yk290b7.3 comes from this gene  
 cDNA EST yk339e8.3 comes from this gene  
 cDNA EST yk471a12.3 comes from this gene  
 cDNA EST yk290b7.5 comes from this gene  
 cDNA EST yk339e8.5 comes from this gene  
 cDNA EST yk471a12.5 comes from this gene"  
 /codon\_start=1  
 /protein\_id="CAB55163.1"  
 /db\_xref="GI:5832826"  
 /db\_xref="SPTREMBL:Q9U2W2"  
 /translation="MKPQWIGLVFVIGIVNCLISSOKLTVEKNEERLKSQGNFTLPIPTAENQNGKGFENDKTYYTNSKSGWLLAAGKRNSTRYATAPFSEHHIILSSRWLTAEHWIMNGKPLDKNNCSGGHHLDVDPVDLDNLIITWGGKPKKADISKAMFRACDNQDLKAPVALIELDKTSKRPMAFNGVCPADLETSSKLDVHVSYGYESQHQRLSTYAKBERICDTDTYETKEKAGGPIVLNVSQKAVIGIKAPASRNELEFYDLKLEICDYGSGCHENPLITTEPPTVETTERSTVSTTSTLETITIMQASSASPSISKPATTEPSPLOASSAPTEPILOPESAPASSASASKEPVOTITPEIOTFSLPTSSLSKTSNTVTTTYTTTTPDPTVDYGDHDHMITGLVDEDEDETTESNNRAHRMDLILGOFEPFLSFFTETCQKI"  
 join(29644..29686,29735..30195)  
 /gene="Y116A8A.7"  
 join(29644..29686,29735..30195)  
 /gene="Y116A8A.7"  
 /note="contains similarity to Pfam domain: PF00635 (MSP (Major sperm protein) domain), Score=61.1, E-value=7.5e-15, N=1"  
 /codon\_start=1  
 /protein\_id="CAB55166.2"  
 /db\_xref="GI:6425238"  
 /db\_xref="SPTREMBL:Q9U2V8"  
 /translation="MNNFLSLCLRLLEDFHDLFSTKNNDSAFIVLSITKRNITPEYOSVPPGVATIEPAKKIIFENAPFDSMTHTYQKAVNLNRTTAYNITLNMKRFSTPPCGILKKTQNMFEITFAPFNTTENTKNDRTITVAKWIMPNNEDDEYFREMFGGQVNHQHHIYOYNI"  
 join(31256..31363,31418..31531,31575..31728,31777..31947,31990..32269,32322..32376)  
 /gene="Y116A8A.8"  
 join(31256..31363,31418..31531,31575..31728,31777..31947,31990..32269,32322..32376)  
 /gene="Y116A8A.8"  
 /note="predicted using GeneParser  
 cDNA EST yk540c6.5 comes from this gene"  
 /codon\_start=1  
 /protein\_id="CAB55160.1"  
 /db\_xref="GI:5832823"  
 /db\_xref="SPTREMBL:Q9U2W5"  
 /translation="MITHHRLALMTCNSLEYSTTETWRCIELCAHNSCVIAQTNLTVEKWCSTQNLNMYTQKRFSANNMMIILAKTDTINETIESVCTPEQGIIVGNFNNTSNILOKYNASQPEEDGNPLHFESYSAIKSQSEBEPDRIRGCMCIKVFQSGSHDAHRACKENAVATLTETREMEYVWNTAYVEALIKQKSMVEYLETWIDGVRNCEQVPAWKSAGCGLDGEFEDEFEFISKGGYTSMDKNPRLDYNQCVLIWIMPNKKEVDDACGSNKKINGFGACKEAG"  
 complement(join(47919..48041,49593..50165,50969..51125,51584..51908,52795..52899,54267..54318))  
 /gene="Y116A8A.9"  
 complement(join(47919..48041,49593..50165,50969..51125,51584..51908,52795..52899,54267..54318))  
 /gene="Y116A8A.9"  
 /note="contains similarity to Pfam domain: PF00557 (metallopeptidase family M24), Score=239.2, E-value=1.9e-68, N=1  
 cDNA EST EMBL:Z14814 comes from this gene  
 cDNA EST EMBL:Z14913 comes from this gene  
 cDNA EST EMBL:M89428 comes from this gene  
 cDNA EST yk138a12.5 comes from this gene  
 cDNA EST yk216d3.5 comes from this gene  
 cDNA EST yk284a3.5 comes from this gene

[illegible]

Fragment Name Begin End  
 CEY113B8\_0 1 110000  
 CEY113B8\_1 100001 210000  
 CEY113B8\_2 200001 310000  
 CEY113B8\_3 300001 357994  
 Continuation (3 of 4) of CEY113B8 from base 200001 (A1021573 Caenorhabditis elegans chr8)

Query Match 12.5%; Score 39.4; DB 2; Length 110000;  
 Best Local Similarity 47.1%; Pred. No. 11;  
 Matches 121; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

Qy 12 tcaagatttaagaataaattttgagaataagaacaataagaatcagtttccaccac 71  
 Db 2826 TCATGCTTTTAATGACATTTTTCCTTAATGATGAATGGCCCTTGATTTTAAATAAA 2767  
 Qy 72 taagatactcttagagatctacagcctcccttaaggagacacacaaagtcagttgt 131  
 Db 2766 TGTATATGATATATTTATTCACAGAACACAGATGAATTTTCTAAMAACTTCACCTTTCA 2707  
 Qy 132 tgccttgtagtccaccctatatatcaagtagtactacaattttgagaataaga 191  
 Db 2706 TCTTTGACTCAATTTTGCACATGCTTACACTTTTGAGTCATATATGATTTGTA 2647  
 Qy 192 ttgtcaccaataaactcggagtttatggaacacatcagtagaagaataacacattccat 251  
 Db 2646 AAGTCAAAAAATTAACCTCAATTAATGATGATGATGACACAGGAAATCTTAAGAAATG 2587  
 Qy 252 ccccttaagaagatcat 268  
 Db 2586 AAATTTGCAAAAATCAT 2570

RESULT 10  
 AC092859/c 171179 bp DNA linear HTG 01-NOV-2001  
 LOCUS Pan troglodytes clone rp43-11lm15, WORKING DRAFT SEQUENCE, 8  
 DEFINITION unordered pieces.

AC092859 AC092859 GI:16519522  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE chimpanzee.  
 ORGANISM Pan troglodytes

REFERENCE AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 TITLE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.  
 JOURNAL 1 (bases 1 to 171179)

REFERENCE AUTHORS N.Y., Song, L. and Roe, B.A.  
 TITLE Pan troglodytes BAC Clone rp43-11lm15  
 JOURNAL 2 (bases 1 to 171179)  
 DIRECT SUBMISSION Submitted (31-JUL-2001) Department of Chemistry And Biochemistry,  
 The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
 OK 73019, USA  
 COMMENT On Oct 30, 2001 this sequence version replaced gi:16445160.

Center: Department of Chemistry And Biochemistry  
 The University of Oklahoma  
 Center code:UOKNOR

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 8 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 2756: contig of 2756 bp in length  
 \* 2757 2856: gap of unknown length  
 \* 2857 5886: contig of 3030 bp in length  
 \* 5887 5986: gap of unknown length

\* 5987 12277: contig of 6291 bp in length  
 \* 12278 12377: gap of unknown length  
 \* 12378 20375: contig of 7998 bp in length  
 \* 20375 20476: gap of unknown length  
 \* 20476 37433: contig of 16957 bp in length  
 \* 37433 37532: gap of unknown length  
 \* 37532 60493: contig of 22960 bp in length  
 \* 60493 60592: gap of unknown length  
 \* 60592 108314: contig of 47721 bp in length  
 \* 108314 171179: gap of 62766 bp in length.  
 \* 108414 location/Qualifiers  
 1. 171179

FEATURES  
 source  
 /organism="Pan troglodytes"  
 /db\_xref="taxon:9598"  
 /clone="rp43-11lm15"  
 /clone\_1lb="RPCT-43"  
 BASE COUNT 47434 a 39162 c 37372 g 46496 t 715 others  
 ORIGIN

Query Match 12.5%; Score 39.4; DB 2; Length 171179;  
 Best Local Similarity 46.5%; Pred. No. 9.9;  
 Matches 127; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

Qy 9 gactcagatttaagaataaattttgagaataagaacaataagaatcagtttccacc 68  
 Db 27477 GATTTCAGAGCTTCTGTAAGTAATGATGAAGAAAAAGCAATTAATCTTACCTTGATGACC 27418  
 Qy 69 acttaagatatactcttagagatctacagcctcccttaaggagacatacaagtcagttg 128  
 Db 27417 ACTGTAGAAATTTAATTAAGCTATCATCACTTGCACACTAAGATGTATTAAGGTAGGCA 27358  
 Qy 129 tttgtccttgtagtccaccctatatatcaagtagtactacaattttgagaata 188  
 Db 27357 AGAATCCTGAAACGTTCTTAAGCCCAAGTCTCATTTGAAGAAAGATAAGCAAGTTTCCAAAT 27298  
 Qy 189 agattgtcacacaataaactcggagtttatggaacacatcagtagaagaagaatacaacattc 248  
 Db 27297 ACAGTCATATGCTGCAATTAACATGTTTGTGTCACACAGAGGCTGCAATATATGCTGAG 27238  
 Qy 249 catcccttaagaagatcatattacttgcaattc 281  
 Db 27237 TGATCCTGTAAGATTTAATGAGCTGAACAC 27205

RESULT 11  
 HSJ364H10/c 150355 bp DNA linear PRI 21-FEB-2000  
 LOCUS Human DNA sequence from clone RP3-364H10 on chromosome 6q16.1-16.3  
 DEFINITION Contains STSS and GSSs, complete sequence.  
 ACCESSION AL078603  
 VERSION AL078603.4 GI:6456827  
 KEYWORDS HTG.

SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 TITLE 1 (bases 1 to 150355)  
 JOURNAL Williams, S.

REFERENCE AUTHORS Direct Submission  
 TITLE Submitted (21-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 REQUESTS: clonerequests@sanger.ac.uk  
 COMMENT On Nov 21, 1999 this sequence version replaced gi:5738453.

\* During sequence assembly data is compared from overlapping clones.  
 \* Where differences are found these are annotated as variations  
 \* together with a note of the overlapping clone name. Note that the  
 \* variation annotation may not be found in the sequence submission  
 \* corresponding to the overlapping clone, as we submit sequences with  
 \* only a small overlap as described above.  
 \* This sequence has been finished according to sequence map criteria  
 \* as follows. An attempt is made to resolve all sequencing problems,



such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at

<http://www.sanger.ac.uk/Projects/C.elegans/wormpep> RPi-364H10 is Institute by the library RPi-3 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/> VECTOR: pCYPAC2

IMPORTANT: This sequence is not the entire insert of clone RPi-364H10. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true right end of clone RPi-364H10 is at 150355 in this sequence. The true right end of clone RPi-273N12 is at 100 in this sequence. This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6>.

## FEATURES

## source

1. 150355

Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="q16.1-16.3"  
/clone="RPi-364H10"  
/clone\_1fb="RPi-3"  
3181..3328  
/note="4 copies 12 mer 85% conserved"  
repeat\_region  
3183..3326  
/note="22 copies 2 mer ca 88% conserved"  
repeat\_region  
7440..7511  
/note="6 copies 12 mer 87% conserved"  
repeat\_region  
7441..7508  
/note="34 copies 2 mer ag 89% conserved"  
misc\_feature  
complement(8514..9064)  
/note="match: GSS: Em:AQ317776"  
9398..9517  
/note="10 copies 12 mer 64% conserved"  
repeat\_region  
9431..9515  
/note="17 copies 5 mer atata 65% conserved"  
repeat\_region  
9875..10098  
/note="HERVL repeat: matches 4335..4548 of consensus"  
repeat\_region  
10429..10684  
/note="MER77 repeat: matches 81..336 of consensus"  
repeat\_region  
10740..11170  
/note="MSTD repeat: matches 1..394 of consensus"  
repeat\_region  
15206..15773  
/note="MER89 repeat: matches 1..559 of consensus"  
misc\_feature  
complement(16716..17027)  
/note="match: STS: Em:HSPA32D12"  
21089..21629  
/note="match: STS: Em:Z78974"  
21100..21488  
/note="match: GSS: Em:AQ030832"  
23485..23606  
/note="61 copies 2 mer ac 64% conserved"  
repeat\_region  
23490..23609  
/note="10 copies 12 mer 69% conserved"  
misc\_feature  
23972..24389  
/note="match: GSS: Em:AQ240903"  
24019..24541  
/note="match: GSS: Em:AQ310408"  
28622..28851  
/note="match: GSS: Em:B75442"  
misc\_feature  
complement(29327..29567)  
/note="match: GSS: Em:AQ279883"  
29568..30147  
/note="match: GSS: Em:AQ322861"

repeat\_region  
30996..31342  
/note="MER58B repeat: matches 2..340 of consensus"  
misc\_feature  
complement(40152..40587)  
/note="match: GSS: Em:AQ177544"  
misc\_feature  
complement(40213..40811)  
/note="match: GSS: Em:AQ014036; match: STS: Em:G51620"  
misc\_feature  
complement(40357..40781)  
/note="match: GSS: Em:B86714"  
40819..41341  
/note="match: GSS: Em:AQ211747"  
46207..46457  
/note="match: GSS: Em:AQ307766"  
49669..49929  
/note="match: GSS: Em:AQ307766"  
repeat\_region  
52044..52403  
/note="L1M4A repeat: matches 6036..6300 of consensus"  
misc\_feature  
53699..54781  
/note="match: GSS: Em:AQ165687"  
repeat\_region  
54775..54811  
/note="Tig1ger3(Golem) repeat: matches 1..1054 of consensus"  
repeat\_region  
54819..54966  
/note="Tig1ger3(Golem) repeat: matches 2298..2456 of consensus"  
repeat\_region  
54957..55129  
/note="Tig1ger3b repeat: matches 1069..1241 of consensus"  
repeat\_region  
61258..61598  
/note="L1R18A repeat: matches 1..346 of consensus"  
repeat\_region  
61600..62465  
/note="HERVL18 repeat: matches 5189..6047 of consensus"  
repeat\_region  
62466..62921  
/note="MUT2B repeat: matches 4..444 of consensus"  
repeat\_region  
62922..68052  
/note="HERVL18 repeat: matches 5..5189 of consensus"  
repeat\_region  
68053..68404  
/note="L1R18A repeat: matches 1..356 of consensus"  
misc\_feature  
complement(70793..71290)  
/note="match: GSS: Em:AQ008841"  
misc\_feature  
complement(71012..71269)  
/note="match: GSS: Em:B83962"  
misc\_feature  
complement(74346..74831)  
/note="match: STS: Em:Z79249"  
75056..75504  
/note="L1MC1 repeat: matches 5653..6108 of consensus"  
repeat\_region  
75929..76148  
/note="L1MC1 repeat: matches 6104..6327 of consensus"  
repeat\_region  
78209..78430  
/note="L1R16B repeat: matches 218..462 of consensus"  
repeat\_region  
78873..79225  
/note="MER58B repeat: matches 1..337 of consensus"  
repeat\_region  
84416..86511  
/note="L1P repeat: matches 1601..3701 of consensus"  
misc\_feature  
complement(86523..86971)  
/note="match: GSS: Em:B88696"  
88350..88397  
/note="4 copies 12 mer 100% conserved"  
repeat\_region  
92738..93167  
/note="MUT2B repeat: matches 2..448 of consensus"  
repeat\_region  
93344..93385  
/note="MUT2B repeat: matches 459..504 of consensus"  
repeat\_region  
97891..98411  
/note="MUT2B repeat: matches 1..448 of consensus"  
repeat\_region  
100416..100639  
/note="MER63A repeat: matches 2..206 of consensus"  
misc\_feature  
complement(101069..101516)  
/note="match: GSS: Em:AQ211414"  
102301..103031  
/note="match: GSS: Em:AQ027024"  
misc\_feature  
complement(109215..109848)  
/note="match: GSS: Em:B56601"  
misc\_feature  
complement(109418..109858)  
/note="match: GSS: Em:AQ217743"  
misc\_feature  
complement(109446..109864)





```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: H_NH0748B21
----- Summary Statistics -----
Sequencing vector: M13, 48x
Sequencing vector: plasmid, 48x
Chemistry: Dye-Primer ET; 48% of reads
Chemistry: Dye-Terminator Big Dye; 50% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 217808 bases at least Q40
Consensus quality: 218152 bases at least Q20
Consensus quality: 218644 bases at least Q20
Insert size: 211000; agarose-1p
Insert size: 218739; sum-of-contigs
Quality coverage: 8.54 in Q20 bases; agarose-1p
Quality coverage: 8.23 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 24044: contig of 24044 bp in length
* 24045 24144: gap of unknown length
* 24145 84338: contig of 60194 bp in length
* 84339 84438: gap of unknown length
* 84439 218939: contig of 134501 bp in length.
*
Location/Qualifiers
1. 218939
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-748B21"
1. 24044
/note="assembly_name:Contig22"
24145..84338
/note="assembly_name:Contig23
clone_end:17
vector_side:left"
84439..218939
/note="assembly_name:Contig24
clone_end:5P6
vector_side:right"
-----
BASE COUNT 73347 a 38196 c 37712 g 69483 t 201 others
ORIGIN
Query Match 12.4% Score 39; DB 2; Length 218939;
Best Local Similarity 54.5%; Pred. NO. 12;
Matches 78; Conservative 0; Mismatches 65; Indels 0; Gaps 0;
QY 119 aagtcagtggtgtgcttgcttgtagtcaccacattatctcaagtagtagtgaactacaa 178
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 134613 AAGAGAGGGTGGAGAGCATTTACTTCAACCACTTAAGACAAATTTCTTATGTGTCACA 134672
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 179 tttagaataatagatgtcacacaataaaccctgaggttatggaacacatcagtagaagaa 238
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 134673 CCATGAAGACAGATAGTCTACCTGAGACTAAAGAACAAATGAAGAACATTCATTAATGAGA 134732
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 239 tacacatccatccoccttaacag 261
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 134733 TTCAACCAATTTAACCCTTTTAAAG 134755
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: September 16, 2002, 02:57:19  
Job time: 7652 sec

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 00:49:47 : Search time 3785.3 Seconds  
(without alignments)  
9287.661 Million cell updates/sec

Title: US-09-700-770-2

Perfect score: 1680

Sequence: 1 gctgctgagagataaagtt.....tcaataaacacttcctctgt 1680

Scoring table:

IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb.ba:\*  
2: gb.htg:\*  
3: gb.in:\*  
4: gb.om:\*  
5: gb.ov:\*  
6: gb.pat:\*  
7: gb.ph:\*  
8: gb.pl:\*  
9: gb.pr:\*  
10: gb.ro:\*  
11: gb.sts:\*  
12: gb.sy:\*  
13: gb.un:\*  
14: gb.vl:\*  
15: em.ba:\*  
16: em.fun:\*  
17: em.hum:\*  
18: em.in:\*  
19: em.mu:\*  
20: em.om:\*  
21: em.or:\*  
22: em.ov:\*  
23: em.pat:\*  
24: em.ph:\*  
25: em.pl:\*  
26: em.ro:\*  
27: em.sts:\*  
28: em.un:\*  
29: em.vl:\*  
30: em.htg.hum:\*  
31: em.htg.inv:\*  
32: em.htg.other:\*  
33: em.htgo.inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
------------	-------	-------	-------	--------	----	----	-------------

1	1679.2	100.0	1680	6	AX224643	AX224643 Sequence
2	1656.2	98.6	2036	6	AX224647	AX224647 Sequence
3	1617	96.2	1636	6	AX092346	AX092346 Sequence
4	1615.8	96.2	1684	9	BC008429	BC008429 Homo sapi
5	822.8	49.0	1629	10	MMU46068	U46068 Mus musculu
6	806	48.0	1035	6	AX301908	AX301908 Sequence
7	724.2	43.1	963	6	AX301906	AX301906 Sequence
8	181.4	10.8	159440	9	AL355392	AL355392 Human DNA
9	161	9.6	161593	2	HSRA49G10	AL121901 Human DNA
c 10	83.2	5.0	89536	2	AC094188	AC094188 Rattus no
c 11	80.6	4.8	92996	2	AC094264	AC094264 Rattus no
12	77.6	4.6	68591	2	AC107771	AC107771 Mus muscu
13	59.4	3.5	89536	2	AC094188	AC094188 Rattus no
14	58.8	3.2	125020	9	AF429315	AF429315 Homo sapi
15	54.6	3.2	999	4	BTU79413	U79413 Bos taurus
16	54	3.2	92996	2	AC094264	AC094264 Rattus no
17	50.4	3.0	1011	9	AF172993	AF172993 Homo sapi
18	48.8	2.9	768	6	E53840	E53840 LUNX gene a
19	48.8	2.9	1015	6	E53841	E53841 LUNX gene a
20	48.8	2.9	1015	9	AB024937	AB024937 Homo sapi
21	48.8	2.9	1053	9	BC012549	BC012549 Homo sapi
22	48.8	2.9	1084	9	AF158745	AF158745 Homo sapi
23	47.8	2.8	33285	1	SC5F8	AL357613 Streptomy
24	47	2.8	5676	9	AB033050	AB033050 Homo sapi
c 25	46.8	2.8	185005	2	AC094500	AC094500 Rattus no
26	46	2.7	1696	4	AB055894	AB055894 Bos tauru
27	46	2.7	31382	1	SC3C3	AL031231 Streptomy
28	45.2	2.7	1500	1	AVSUCCTR	X52432 A.vinelandi
29	45	2.7	3312	1	AF061246	AF061246 Ralstonia
30	44.8	2.7	1561	10	AF194029	AF194029 Rattus no
c 31	44.8	2.7	125020	9	AF429315	AF429315 Homo sapi
c 32	44.6	2.7	1692	3	AF350278	AF350278 Nephtila m
33	44.4	2.6	336	9	AF369213	AF369213 Homo sapi
c 34	44.2	2.6	45335	2	SCC57A	AL136319 Streptomy
c 35	44.2	2.6	68591	2	AC107771	AC107771 Mus muscu
36	43.8	2.6	1675	10	AB055895	AB055895 Rattus no
37	43.8	2.6	1713	10	AB055896	AB055896 Rattus no
c 38	43.8	2.6	12213	1	AE004441	AE004441 Pseudomon
39	43.4	2.6	20102	1	SC9B2	AL055212 Streptomy
c 40	43	2.6	174384	2	AC068495	AC068495 Mus muscu
41	42.8	2.5	1668	10	AF194028	AF194028 Mus muscu
42	42.8	2.5	1691	10	BC011168	BC011168 Mus muscu
43	42.8	2.5	1709	10	AB055897	AB055897 Mus muscu
44	42.8	2.5	13591	10	AB055898	AB055898 Mus muscu
45	42.6	2.5	1584	5	GGRRG2	X73973 G.gallus RA

## ALIGNMENTS

RESULT	1
AX224643	
LOCUS	AX224643 1680 bp DNA linear PAT 10-SEP-2001
DEFINITION	Sequence 1 from Patent WO0161055.
ACCESSION	AX224643
VERSION	AX224643.1 GI:15554772
KEYWORDS	
SOURCE	human.
ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
REFERENCE	1 (bases 1 to 1680)
AUTHORS	Chen,S.Y., Sun,Y. and Macina,R.A.
TITLE	Methods for diagnosing, monitoring, staging, imaging and treating
JOURNAL	lung cancer via lung cancer specific genes
diabexus, Inc. (US)	Patent: WO 0161055-A 1 23-AUG-2001;
FEATURES	Location/Qualifiers
SOURCE	1..1680
BASE COUNT	380 a 503 c 429 g 367 t 1 others
ORIGIN	

Query Match	98.6%	Score 1656.2	DB 6	Length 2036
Best Local Similarity	99.7%	Pred. No. 0		



Matches 1678; Conservative 2; Mismatches 0; Indels 3; Gaps 2;

```
QY 1 ggtgtcagagataaagtgtgacttcacagaccactgcgggagag--gagrgagcg 58
Db 38 GGTGTCAAGATTAAGTTGGACTTCCAGACCCACTGCCCGGAGAGAGAGAGAGCG 97
QY 59 ggcgcgagatccacagctgcccaggtctgcatctctgcatctgtgccccttgacactg 118
Db 98 GGC CGAGSACTCCAGGCTGCCAGGTCTGGCATCTCGACTTGTCTCCCTCGACACCTG 157
QY 119 ggaagatgacgagcccggtgacctcaacctctctgtgtgtgtgtgtgtgtgtgtgt 178
Db 158 GGAAGATGGCCGCGCGTGGACCTTCACCTTCTGTGTGTGTGTGTGTGTGTGTGT 217
QY 179 tcgaagccacctcaagtcacatctgcatctcaatctctgtgtgtgtgtgtgtgtgtgt 238
Db 218 TCGAAGCCACCTCATGCTCCACTGCGAGTTCTCATCTCGGCCCAAAAGTCATCAAGAA 277
QY 239 agctgacacagagagctgaaagacacacacacacacacacacacacacacacacacac 298
Db 278 ACTGACACAGAGAGCTGAAGACACACACACACACACACACACACACACACACACAC 337
QY 299 tcaatgcaatgcgagaaagacacacacacacacacacacacacacacacacacacac 357
Db 338 TCAGTGCATGCGGGGAAAAAGCCAGCCGAGGACATCCCTGTGTGTGTGTGTGTGTGT 397
QY 358 cgtctctgagacacatcatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 417
Db 398 CGGTCTGGAAGACATCATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 457
QY 418 tgaagcccttcgacacacacacacacacacacacacacacacacacacacacacacac 477
Db 458 TGAAGCCCTCGGCAATGACACAGAGCTCTAGTCAAGATCCCTCGGACATGTGTGTGT 517
QY 478 gattcaacacgcccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 537
Db 518 GATTCAACACGCCCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 577
QY 538 ccaccaatccgcatgagacacacacacacacacacacacacacacacacacacacac 597
Db 578 CCACCATCCGCAATGACACACACACACACACACACACACACACACACACACACAC 637
QY 598 ccaccaatccgcatgagacacacacacacacacacacacacacacacacacacacac 657
Db 638 CCACCATCCGCAATGAGAGCTCGGCATCCAGTGTGTGTGTGTGTGTGTGTGTGTGT 697
QY 658 ccttagcttagacagatcatgaacctctctgtgtgtgtgtgtgtgtgtgtgtgtgt 717
Db 698 CCTTAGCTAAGCAGGTCATGAACCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 757
QY 718 agctgtgtccgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 777
Db 758 AGCTGTGTCCGTCATGAGGCTCTCTTAATGTGTGTGTGTGTGTGTGTGTGTGTGT 817
QY 778 tgaaggttcacatccctcagacatgacacacacacacacacacacacacacacacac 837
Db 818 TGAAGTGTCCGTCATGAGGCTCTCTTAATGTGTGTGTGTGTGTGTGTGTGTGTGT 877
QY 838 tcaaggtgacacacacacacacacacacacacacacacacacacacacacacacac 897
Db 878 TCAAGGTGTACACACATTCAGCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 937
QY 898 tgaacagatggttcaatactctgacacacacacacacacacacacacacacacacac 957
Db 938 TGAACAGATGTTCAATTAATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 997
QY 958 cgttcagacctcatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1017
Db 998 CGTTCAAGCTCATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1057
QY 1018 caagaagaatcatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1077
Db 1058 CAGAAGAATTCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1117
```

```
QY 1078 caagcatcggtctgatalcaatgaagaagctgcagataaagtgtggtatctaccagatcgtga 1137
Db 1118 CAAGCATCGGCTGATCATTAAGTAAGAGCTGCGATTAAGCTGGATATACCCAGATGTGA 1177
QY 1138 agatccatacagacacacacacacacacacacacacacacacacacacacacacacac 1197
Db 1178 AGATCTTAATCAAGACACATCCGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1237
QY 1198 aactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1257
Db 1238 AACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1297
QY 1258 gcatcgaagccagctgcggaagctcagtttatacacaagatgacacacacacacacac 1317
Db 1298 GCATCGAACCGAGCTGGAAGCTCACTTTTACACCAAGGTGACCACTTAATCACT 1357
QY 1318 tgaatatacagctctgatalcagatcagctgatacactgtggtatgtgtgtgtgtgtgt 1377
Db 1358 TGAATTAACATCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1417
QY 1378 ctgagtgtctgaaacacatcatcactgagatcatcactccatctgtgtgtgtgtgtgt 1437
Db 1418 CTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1477
QY 1438 atgagcaaatgaatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1497
Db 1478 ATGGCAATTAAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1537
QY 1498 ctgagtcctcactgacacacacacacacacacacacacacacacacacacacacacac 1557
Db 1538 CTGAGTCTCTACATGACCAAGATGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1597
QY 1558 gctctctgtctccacagtgaaagactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1617
Db 1598 GCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1657
QY 1618 tggagatgagtgatgagctctataagacatccctctctgcaatcaataaacactgtgt 1677
Db 1658 TGGAGATGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1717
QY 1678 gttg 1680
Db 1718 GTG 1720
```

RESULT 3  
AX092346 1636 bp DNA linear PAT 21-MAR-2001  
LOCUS AX092346  
DEFINITION Sequence 77 from Patent WO0116318.  
ACCESSION AX092346  
VERSION AX092346.1 GI:13444487  
KEYWORDS  
SOURCE  
ORGANISM human.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1636)  
Eaton,D.L., Filvaroff,E., Gerritsen,M.E., Goddard,A.,  
Godowski,P.J., Grimaldi,C.J., Gurney,A.L., Watanabe,C.K. and  
Wood,M.I.  
TITLE Secreted and transmembrane polypeptides and nucleic acids encoding  
the same  
JOURNAL Patent: WO 0116318-A 77 08-MAR-2001;  
Genentech, Inc. (US)  
FEATURES  
source Location/Qualifiers  
1..1636  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 374 a 492 c 411 g 359 t  
ORIGIN

Query Match	96.2%	Score 1617	DB 6	Length 1636
Best Local Similarity	99.8%	Pred. No. 0		
Matches 1628	Conservative 1	Mismatches 1	Indels 1	Gaps 1
QY 51	grrgagcggccgagagaccacacgctgacgagctgctgagatccctgcaacttgcctcctc	110		
DB 1	GAAGAGCGGGCCCGAGACTCCAGCGGCCAGGTCGGCATCCTGCACACTTGGCTGCCCTCT	60		
QY 111	gaacacgtggaagatgagccgagccgtgagaccccttcctctgtgtgttctgagac	170		
DB 61	GAACACTGGGAAGATGGCGCGCCCGTGGACCTTTCCTCTGTGGTTCTTGCGAC	120		
QY 171	caacttgatccaaagcaccctcaagtcacatgcttcacatccctcgcgcaaaagtc	230		
DB 121	CACCTTGATCCAGACCCCTCAGTCCACTGCTTCATCTCTCGGCGCAAAAGTCA	180		
QY 231	caaaagaagaagtgaacagagctggaagacacacacgacacacacacacacacac	290		
DB 181	CAAGAAGAAAGTGAACAGAGACTGAAGACCAACACGACACACATCTGCACACT	240		
QY 291	gccgctgtcagtgccatgagggaaagccagccgagag-alcctgtgtgtgagcct	349		
DB 241	GGCGGTGCTCAGTGGCATGCGGAAAGCGACCGAGGCAATCCCTGTGGTGGCAGCT	300		
QY 350	gtggaacacccgtccgaagacacatctctgtgtgaagtcacacagtaacatccca	409		
DB 301	GGTGAACACCGCTCTGAAGACATCATCTGGTGAAGGTGATCAACAGATTCCTCA	360		
QY 410	gctgagagtgaaagccctcgacacatgacagagagctgtgagtaagatcccccctgag	469		
DB 361	GCTGAGGTGAAGCCCTCGGCAATGACAGAGAGTGTGATCAAGATCCCTCGACAT	420		
QY 470	gttgagctgagttcaacagccctgtgtcaagacatgltgagttcaacatgagcagta	529		
DB 421	GCTGCTGATTCACACGCGCCCTGTCTCAAGACCATCTGGAAGTTCACATGACAGCA	480		
QY 530	ggcccaagccacacatccgacatggaacacagtgacagtgagcccccgcgctgtcag	589		
DB 481	GGCCCAAGCCACATCCGATGAGACAGACAGTGAAGTGGCCCCCGCGCTGGTCTAG	540		
QY 590	tgactgtgcaacagcagatggagcctgagcatalccaaactgtctgataagctctcct	649		
DB 541	TGACTGTGCCACAGCCATGGAGCCGTCGCATCCAACTGTGTATGAAGCTCTCTCT	600		
QY 650	gttgaacgcttaagcagatgacagatgacatctctgtgtgacatccctgcccacatg	709		
DB 601	GGTGAACCCCTTACCTAAGCAGGTGATGACCTCTAGTGCATCCCTGCCAATCTAGT	660		
QY 710	gaaacacagctgtgtccctgagcagagcttccctcaatgagcagatgacagacccct	769		
DB 661	GAAACACAGCTGTGCTCCGATGAGGGCTCTCTCAATGGCATGTATGACAGACTCT	720		
QY 770	gcagctgtggaaggtgcccattccctcagcaatgacagctgtgagttgacattctgta	829		
DB 721	GCAGCTGTGAAGGTGCCATTTCCTCAGCATTTGACCGTGTGACCTTCTGTA	780		
QY 830	tcctgcatcaaggtgtaacacattagctctacatgaggggacaaagtgtgtgaccca	889		
DB 781	TCCGTCATCAAGGGTACACATTCAGCTCTACCTGCGGAGCAAGTTGTTGAGCTACA	840		
QY 890	gggaaaggtgacaaagtgtgtcaataactgtgacgttccctgtacaaatgcccacctgga	949		
DB 841	GGGAAAGGTGACCAAGGTGTTCAATCTGCAAGCTTCCTCAACATGCCACCTGGA	900		
QY 950	caacatcccglttagcctcagtgtagtaagagcgtgtgtgaaagctgacagtgagctgt	1009		
DB 901	CAACATCCCGTTACGCTCTACCTGTGATGACGAGTGTGAAAGCTGACAGTGGCTCTGT	960		
QY 1010	gctctctcagaagaatcattagtlcctgtgtgagctgtgtcctcagagatggccatg	1069		
DB 961	GCTCTCTCAGAGAAATTCATGTGCTCTGTGGACTGTGCTTCCGAGAGTGGCCATG	1020		
QY 1070	gctgaagtgcaagatcggtgctgatacaatgaagagctgacagataagctgtgagatacca	1129		

DB 1021	GCTGAAGTCAAGCATCGGGCTGATCAATGAAGAGCTGCACATAGCTGGGATCTACCA	1080		
QY 1130	gacgtgagaatccaaactcagacacccagttttttaaagcaagccatgcca	1189		
DB 1081	GATCGTGAAGATCCTAATCTAGAGACACTCCGAGATTTTATACACAAAGGCCATGCGA	1140		
QY 1190	gttgagccaaactgacgtgtgtggaagtgttccctcagtgaaagccctcgccttgt	1249		
DB 1141	GGTGGCCCACTGATCGCTGGAAGTGTTCCTCCAGTGAAGCCCTCGCCCTTGT	1200		
QY 1250	cacctggagcatcgaagccagctggaagctcaatttaacacaaagttaaccaactat	1309		
DB 1201	CACCTGGGCTTCAGACCCAGCTGGAAGCTCACTTTTACCAAGGTGAACAATTA	1260		
QY 1310	actcaacttgataataactcagctctgacagctgacagctgacagctgagctgtg	1369		
DB 1261	ACTCAACTTGAAATTAATCACTGAGCTGTGATCGGATCCAGCTGATGAACCTGGAT	1320		
QY 1370	gttcaacactgagtttctgaaaaaacatcatcaetgagatcaatcaatccatcctgtgc	1429		
DB 1321	GTTCAACCTGATGTTCTGAAAAACATCATCACTGAGATCATCCATCCTGCTGCC	1380		
QY 1430	gaaccagaatgcaaatgaatcctggtgtccagtgatgltgtgaaagcccttggatt	1489		
DB 1381	GAACCAAGATGGCAATTAATGATCTGGGGTCCAGTGTCTATTGTGAAGGCTTGGGATT	1440		
QY 1490	cgagagcagctgagctcactgacacaaagatgacctgtgtactacacagcctcctgtg	1549		
DB 1441	CGAGGCGAGTGAATCCCTCACTGACCAAGATGCGCTTGTGATTCACAGCTCTTGTG	1500		
QY 1550	gaaacacagctcctcgtgtccagtgaaagacttgatgagcagcctcaggaagctgtg	1609		
DB 1501	GAAACCGAGCTCTCTGTCTCCAGTGAACACTTGGATGGACCAATGAGGAAGCTG	1560		
QY 1610	gtccagctgtggagatggtgtgtgagctctatagacacatccctcctgcaatcaaaac	1669		
DB 1561	GTCACGCTGGGAGTATGGGTGAGCTCATATGACCATCCCTCTGCATCAATTAAC	1620		
QY 1670	acttgctgtg 1680			
DB 1621	ACTTGCTGTG 1631			
RESULT 4				
BC008429	1684 bp	mRNA	linear	PRI 12-JUL-2001
LOCUS				
DEFINITION	Homo sapiens, Similar to DNA segment, Chr 2, Massachusetts Institute of Technology 19, clone MGC:14597 IMAGE:4291561, mRNA, complete cds.			
ACCESSION	BC008429			
VERSION	BC008429.1	GI:14250057		
KEYWORDS	MGC.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	Strausberg, R.			
TITLE	Direct Submission			
JOURNAL	Submitted (25-MAY-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>			
COMMENT	Contact: MGC help desk Email: <a href="mailto:cgaps-remail.nih.gov">cgaps-remail.nih.gov</a> Tissue Procurement: CLONTECH CDNA Library Preparation: CLONTECH Laboratories, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNI) DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305 Web site: <a href="http://www.shgc.stanford.edu">http://www.shgc.stanford.edu</a>			

Contact: (Dickson, Mark) mcd@axil.stanford.edu  
Dickson, M., Schmitz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNM at: <http://image.llnl.gov>  
Series: IRAL Plate: 21 Row: 3 Column: 12  
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein.

## FEATURES

## source

1. 1684

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="MGC:14597 IMAGE:4291561"

/tissue\_type="Skeletal Muscle"

/clone\_11b="NH\_MGC\_81"

/lab\_host="DH10B"

/note="Vector: PDNR-LIB"

77. 1531

/codon\_start=1

/product="Similar to DNA segment, Chr 2, Massachusetts

Institute of Technology 19"

/protein\_id="AAH08429.1"

/db\_xref="GI:14250058"

/translation="MAGPWTFLLGGLAATLLOATLSPYAVLIIGPKYIKELNOEL

KDHNTSTLLOPLLSAMREKRPAGIPVLSVNTLVKHLVTLKXTITNIILOVLRPS

ANDDELIVKIPLDWYAGRTPLVKTIVPEHMTTEKQATIRMTSASGPRILSDCAT

SHGSLRIQLHLKLSFLVNALAKQVNLVPSLPNVLKQCVIASFEGMTADLLQL

VKPPLSLSIDLEFDLYPAIKGDTIQIYLGKILDSOGKTVKWNNSAASLTMPFLD

NIPSLIYSDVVAALVAVLSPEFMYLDVLESAHLKLSIGLINERKADKGS

TOIVKILDDPEFPIIDOGHAKVAOLIVLEVPESEALRPLFTLIEASSEAOPTKG

DOLLINNNISDRIOQMNSGIMGFOPVYLKNIITETIHSILLPNONGKLRGSEVSL

VKALFEAESLSLDALVLTPLASLPKPSVPSQ"

BASE COUNT  
ORIGIN

419 a 493 c 414 g 358 t

## Query Match

Best Local Similarity 99.7%; Score 1615.8; DB 9; Length 1684;

Matches 1628; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

49 gagtgaagcggcgaggaactcagcgtgccagctgtgacatctgcaactgtgacct 108  
111:|||||  
Db 2 GAGAGGAGGGGGCGAGACTCCAGGCTGCCAGCTGTGGCAATCTGCACTTGTCCCT 61  
109 ctgacaacttggagaatgtgcccggccgttggaactcaacctctctgtgtgttgaa 168  
111:|||||  
Db 62 CTGACACTGGGAGATGGCCGGCCGGTGTGACCTTCTCTGTGTGTGTGGCA 121  
169 gccaccttgatccaagccacctcagctccacactgacgttctatctctggcccaaatc 228  
111:|||||  
Db 122 GCCACCTTGATCCAGCCACCTCACTCCAGTGTCTCATCTCGGCCCAAAAGTC 181  
229 atcaaaagaagctgacacagagactgaaagacacacagccacacagatcctctgacag 288  
111:|||||  
Db 182 ATCAAGAAGAAGCTGACACAGAGCTGAAGAGACACACAGCCACAGCAATCTGCAAGC 241  
289 ctgcccgtctcagtgcatgtcggaagaagccagccggaag -atccctgtgtgtgacg 347  
111:|||||  
Db 242 CTGCCCTCTCTCACTGCTCCATGCGGGAAGCCAGCCGAGGCAATCCCTGTGCTGGCAGC 301  
348 ctggtgaagaacgctctgaagcacatcatctgtgtgaagatcatcaagtaatacatctc 407  
111:|||||  
Db 302 CTGGTAAACACCTCTTGAAGACGTCATCTGCTGAAGGTCAATCAAGTAATCATCTTC 361  
408 cagcttgcaagtgaaagccttcggcacaatgacagagactgtctagtcaagatcccccggac 467  
111:|||||  
Db 362 CACCTCAGGTGAAGCCCTCGGCCAATGACACAGAGCTGTGTATCAAGATCCCTCGAGC 421  
468 atgtgtgtgtgtatcaaacagccctctgtcaagaacatcgttgaagttccaatgacact 527  
111:|||||  
Db 422 ATGTTGGCTGTGATTAACAGCCCTGTGTCAAGACCATGTGTGATTCACATGAGAGCT 481

528 gagggcccaagccacatccgcgatgagacaccagtgagggcccaaccgctgtctc 587  
111:|||||  
Db 482 GAGGCCCAAGCCACCATCCGATGAGACACAGTGAAGTGGCCCAACCCCTGTGTCTC 541  
588 agtgaactgtgcccacagccatggagccctgacatccaaactgtgtgataagctcttc 647  
111:|||||  
Db 542 AGTGACTGTGCCACAGCCATGGAGCTCGCATCCCACTGCTGATAGCTCTCTTC 601  
648 ctgttgaaagccttaagtaagcaggtcaatgaactcctctagtgcacatccctgccaatc 707  
111:|||||  
Db 602 CTGTGAAGCCCTTACTAGTGAAGTCAATGAACTCTTACTAGTCAATCCCTGCCAATCTA 661  
708 gtgaaaacacagctgtgtcccgatgcagagctcctctcaatgtgatatgaagacct 767  
111:|||||  
Db 662 GTGAAAACACAGCTGTGTCCCGATGAGAGCTTCTTCAATGAGCATGTATGAGACTTC 721  
768 ctgcaactgtgtgaagtgcccatctccctcaagcatgaaccctgtgagttgacctctg 827  
111:|||||  
Db 722 CTGCACCTGTGTGAAGGTGCCCATTTCCCTAGCATTTGACCGTGTGGAGTTGACCTTC 781  
828 tatctgtcaatcaaggttgacacatcagctctaccctggggccaagtgtgtgactca 887  
111:|||||  
Db 782 TATCTGTGCATCAAGGGTGCACACATTCAGCTCTACTGTGGGGCCAAAGTTGTGACTCA 841  
888 cagggaagaagtgacaaagtgatcaataactctgagagctcctctgacaaagccacact 947  
111:|||||  
Db 842 CAGGGAAGGTGACCAAGTGTTCATTAATCTGTGACGCTTCCCTGCAAAAGCCACCTCG 901  
948 gaaacaatcccgcttcagcctcactcgtgagtcaggaacgtgtgaaagcctgacgtgact 1007  
111:|||||  
Db 902 GAAACATCCCGTTCAGCCCTCATCTGAGCAGAGAGCTGTGAAGCTGTGAGTGGCTGT 961  
1008 gtgtctctcagaagaatcatgtgtctgtgtgactgtgtctcctggaagtgccat 1067  
111:|||||  
Db 962 GTGCTCTCTCCAGAAATATTCATGTCTGTGTGAGCTGTGCTGCTGAGATGGCCAT 1021  
1068 cagctgaagtcaagcctcgggctgcatcaatgaagaagctgtgaaatgaagctggatctacc 1127  
111:|||||  
Db 1022 CGGCTGAAGCAAGCATCGGGCTGATCAATGAAGAAGCTGCAATGAAGCTGGATCTACC 1081  
1128 cagatcgtgagaatccctaactcaggaacatcccgagttttttaagaaccaagccatgac 1187  
111:|||||  
Db 1082 CAGATGTGAAGATCTTAATCTAGAGACACTCCGAGTTTATTAAGCAAGGCCAATGCC 1141  
1188 aaggtggcccaactgatatcgtgtggaagtgttccctcgaagtgaagccctcgaccttg 1247  
111:|||||  
Db 1142 AAGGTGGCCCACTGATCGTGTGGAAGTGTTCCTCCAGTGAAGCCCTCCGCTTTG 1201  
1248 ttcaaccttgagacatgaaagcagctcggaagctcagttttaaccaaaagtgaaccaatt 1307  
111:|||||  
Db 1202 TTACACCTGGGATTCGAGAGCCAGCTCGGAAGCTCAATTTTACCAAAAGTGCACAACTT 1261  
1308 atactcaacttgataatacatcagctctgatacagatcagatgaactgtgagatgtgac 1367  
111:|||||  
Db 1262 ATACTCAACTTGAATTAACATCAGCTCTGATCGATCCAGCTGTATGAACCTGTGGATTGGC 1321  
1368 tggttccaactgtatgttctgaaaaacatcaactgtgagatcatccaactcatcctgtgt 1427  
111:|||||  
Db 1322 TGGTTCCAACTGATGTTCGAAAAACATCATACGAGATGATCACTCAATCTGCTG 1381  
1428 ccgaacacagaatggaatgaatgaatcgtgggtccaggtgtgcatctgtgtgaagccttgga 1487  
111:|||||  
Db 1382 CCGAACCAGATGCAATTAATTAAGATGTGGGGTCCCACTGTGATGGTGAAGGCCCTTGGGA 1441  
1488 ttctgaagcagctgagctcactcaactgacaaagatgacctgtgtcttaactcagctccttg 1547  
111:|||||  
Db 1442 TTGAGGAGAGCTGAGTCTCCTACAGACAGAGAGTCCCTGTGCTTACTCAGGCTCCTTG 1501  
1548 tggaaacccaagctctcctgttctccagtggaagcttgatggaagcactcaaggaagct 1607  
111:|||||  
Db 1502 TGGAAACCCACCTCTCTCTCTCTCCAGTGAAGACTTGATGGCAGCATCAGGGAAGGCT 1561

QY 1608 ggtccacgctggagatagtggtgagctctataagaccatccctctgcaatcaataa 1667  
|||||  
Db 1562 GGGTCCAGCTGGGAGATGGTGTAGCTCTATAGACCATCCCTCTCTCATCAATTA 1621  
|||||  
QY 1668 aactgctctg 1660  
|||||  
Db 1622 ACACCTGCTGTG 1634  
|||||

RESULT 5  
MMU46068 1629 bp mRNA linear ROD 10-AUG-2000  
LOCUS Mus musculus von Ebner minor salivary gland protein mRNA, complete  
DEFINITION cds.  
ACCESSION U46068 GI:9789706  
VERSION 046068.2  
KEYWORDS  
SOURCE house mouse strain-Swiss-Webster.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 1629)  
Snead, M.L., Villanueva, J., Paine, M.L., Lei, Y.P., Zhu, D.H.,  
Lusis, J., Xia, Y.-R. and Yang, J.-N.  
TITLE Submitted (17-JAN-1996) Malcolm L. Snead, Center for Cranio  
Molecular Biology, University of Southern California, 2250 Alcazar  
St, Los Angeles, CA 90033, USA  
COMMENT On Aug 10, 2000 this sequence version replaced gi:1184789.  
FEATURES  
Location/Qualifiers  
1..1629  
/organism="Mus musculus"  
/strain="Swiss-Webster"  
/db\_xref="taxon:10090"  
/chromosome="2"  
/map="2; DZM119n and DZM125n"  
58..1482  
/codon\_start=1  
product="von Ebner minor salivary gland protein"  
protein\_id="AA87381.2"  
/db\_xref="GI:9789707"  
translation="MAGPWITLLCGILGATLVQAVNYPPPAVNLGPEVYOKRLQAL  
KDHATALIOELPLRAMODSGSIPIDSEFVHTVLTITMKVTSANILIQDVOST  
YNELVRIPLDMWAGLNPILKIVEFQSTVOLAIRERKSGPAHLNLSDCSSN  
ESTRLSLHLKLSFVSNLAKNYMNLVPAIPQIVNNHLCFPMIQAFDMDYEDFLRL  
TAPALSPALFEGCLSPALQDSNLTNLAKLDSQARTNMFNNSATSLMETPDG  
APPSLYRQDLVNAIVTTLVKEELVLLFVLPDVARLOMDIKEINAEANKLQPT  
QMLKTFHSPHTVINEGSAARAQSVLVEFPTNPVREFPSIGISAEQAFPEEDN  
RLMLNFMNNSVIERIKIMISDIKLFDEVLKDLTKLTLEYTLIPNENKLRGTIPMSMS  
KALGYERAMSVSRKALKILTPASS"

CDS  
BASE COUNT 411 a 497 c 382 g 339 t  
ORIGIN

Query Match 49.0%; Score 822.8; DB 10; Length 1629;  
Best Local Similarity 71.2%; Pred. No. 8.5e-185;  
Matches 1129; Conservative 0; Mismatches 452; Indels 5; Gaps 3;

QY 336 gtctgagcagctgtgtgaacaccgtctgaagcacatcatctgtgaagtatcaca 395  
|||||  
Db 268 ATACTGACAGCTTGTGTGCACACCGTTTGAATATCATATGATGATGAGTCACTCT 327  
|||||  
QY 336 gctaacatctccagctgtgcaagtgaaagccctcggccaatgacagagctgtgtaag 455  
|||||  
Db 328 GCTTAACATCTCCACAGCTGGATGTGCAGCCTTCGACCTATTAACAGAGAGCTGTGTGAGA 387  
|||||

QY 456 atccccctggaatgtgtgtggaatgaacagccccctgtgtcaagacatcgtgtgagtc 515  
|||||  
Db 388 ATCCCGCTGGACATGTGTGTGTGAGTAAACACACCTGTATCAAGACCATATGTGAGATTCC 447  
|||||

QY 516 cacatgacgactgagcccaagccacacatccgaatggaacacagttgaagtggccacc 575  
|||||  
Db 448 CAATGAGACACCGAGGTCCAAAGCCCTCATCCGGGTGAGAGAGACCAAGAGCCGCCGCC 507  
|||||

QY 576 gcctgtctcctgaatgt 635  
|||||  
Db 508 CACTGGAATCTCAGCGACTGCTCCAGCAATGAGAGACCCCTGCGCTCAGCTGCTTAC 567  
|||||

QY 636 aagctctcctcctgtgtggaagccttgatgaagcaagtgtaaacctctgagtcacac 695  
|||||  
Db 568 AAGCTCTCTTCTGT 627  
|||||

QY 696 ctgccaatctagtgaaacacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 755  
|||||  
Db 628 CTGCCCCAATATGTAAGAAAACCACTATGCCCCATGATCAGCAGGCTTTGATGATG 687  
|||||

QY 756 tatgaagactcctgtgagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 815  
|||||  
Db 688 TAGGAAGACTTCTGTGAGCTGTGACACACCAACCATTTGCCCTGAGCCCTTGGAG 747  
|||||

QY 816 ttggaactctgtatctcgtccatcaaggttgaaacacatcaactctgacgtctccatga 875  
|||||  
Db 748 TTGGCCCTCTGTCTCTCTGTATCCAGACAGTAATCTCTGTGAATGTGAAGGCCAAG 807  
|||||

QY 876 ttgtgagctcaaggaaggtgtgaacaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 935  
|||||  
Db 808 CTCTGTGACTCACAGGCGAGGGTAACCACTGTTCACAACTGCGCAGCTCTGTGATG 867  
|||||

QY 936 atgcccaccttgagacacatccctgtcagcctcatcgtgtgaagcgtgtgtgtgtgtgt 995  
|||||  
Db 868 GAGACACACCCAGAGCGGGCCCCCTTACACCTGACCGGTGAGGACGACTGTGTGAATGCC 927  
|||||

QY 996 gcaagtgtcgt 1055  
|||||  
Db 928 ATTGTGACACCTGTGCTCCCAAGAGAGAGCTTTGAATCTGCTCAGATTCGTGATTCCT 987  
|||||

QY 1056 gtagatgcccacatcggtgtgaagcagcatcggtgtgtgtgtgtgtgtgtgtgtgtgtgt 1115  
|||||  
Db 988 GATGTGGCCCGCCAGTTCACAGATGACATCAAGAAATCAATGACAGGACGACCAACAG 1047  
|||||

QY 1116 ctggaatctacacagatcgtgaagatccctaaactaagcaactcccgagttttatagac 1175  
|||||  
Db 1048 CTGGGGCCACCCACAGATGTTGAAGATCTTCAACCAACAGACCCCCACATGTGCTGAT 1107  
|||||

QY 1176 caagccatgtgcaaggtgtgccaactgtatgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1235  
|||||  
Db 1108 GAGGCAATGCGCAGGAGGACGCCAAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1167  
|||||

QY 1236 ctgcgcccctgttcaacccctgtgcatcgaagcagcctcgggaagtgtgtgtgtgtgtgt 1295  
|||||  
Db 1168 GTCCGGCCCTTCTCTCTCTGT 1227  
|||||

QY 1296 gtgtgaacatctatcaactgtgaataacatcaagctgtgactgtgacatcgtgtgtgtgt 1355  
|||||  
Db 1228 GACAACCGGCTTTATGCTCAACTTCAATATATGTCTGATGAGGACATCAAGCTGATGATC 1287  
|||||

QY 1356 tctgtgagttgt 1415  
|||||  
Db 1288 TCAATATATCAAACTATTTGATCTCTGAAGTCTGTAAGAGACACCCCTGACCAAGATCTTGA 1347  
|||||

QY	1416	tcacatctcgagccgaacacgaatggaataatgatttgggtgccagtgtaattggtg	1475
Db	1348	TACACATGCTGCGCCAAATGAGAAATGCAAACTGCAACTGGAAATCCCATGATCAATGCTCC	1407
QY	1476	aaggccttgagatctcgagagcctgagatgctccctcaactgacgaagaatgacctgtgttact	1535
Db	1408	AAGGCTTGAGGATACGAGAAAGCCATGTCGTGTGACCAAGGGTGGCCCTCAAGTCACT	1467
QY	1536	ccagctctctgtgtggaacaccagactctcc-tgtctccagtgaagaacttgatgtgcaagc	1594
Db	1468	CCAGCTCTCTCTGTGAAACCTCTGGCTCCCCGGCCCTCTCGATGAAGAACCTCGACAGATGACA	1527
QY	1595	atcaggaagagctggtgtgtccagctggtgagatgtgtgtgagctctatggaaccctcct	1654
Db	1528	GCCAGCCCCCGCCAGAGTCCACGACGAGAGTGTGAAGCTCTGTGTAGAC-TACCCCC	1585
QY	1655	ctgcaatcataataaacactgtgctgtg 1680	
Db	1586	CTGCAGTTAATTAACACTTGGCCCGTG 1611	
RESULT	6		
AX301908	AX301908	1035 bp	DNA
LOCUS	Sequence 15 from Patent WO0174851.		Linear
DEFINITION	AX301908		PAT 30-NOV-2001
ACCESSION	AX301908.1	GI:17382958	
VERSION			
KEYWORDS	human.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE			
AUTHORS	Mejunder,K., Spaderna,S.K., Taupier,R.J., Padigar,M.,		
	Burgess,C.E., Shinkets,R.A., Spytek,K.A., Liu,X., Patturajan,M. and		
	Gusev,V.Y.		
TITLE	Novel proteins and nucleic acids encoding same		
JOURNAL	Patent: WO 0174851-A 15 11-OCT-2001;		
FEATURES	Curagen Corporation (US)		
source	Location/Qualifiers		
	1..1035		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
CDS	79..1035		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="CAD13090.1"		
	/db_xref="GI:17382959"		
	/translation="MAGPWFLLGLLAATLLIOATLSPRAVILILGKVIKERTIOEL		
	KDHATSILOOLPLLSAMREKPGICIVLSLVNTVLAHITLWLVATIANILQDVKFS		
	ANODELVKIPLDVAVGFNPPLVKTIVIEFHMTTEAQTIRNDTSASPTRIVLVDCATN		
	SHGSLRIQLHKLSFLVNALAKQYMNLLVPSLPIVKNQCLPYEASFNGVYADLDLD		
	VKSCSALSPTEFTTSLASRPKVKTKWPNNSASLIMPTLDINIPFSLIVSQDVKAA		
	VAAVLSEEMVILDSVYNLSTRRIORIPRPHRNFINTGCP"		
BASE COUNT	230 a 333 c 261 g 211 t		
ORIGIN			
Query Match	48.0%: Score 806; DB 6; Length 1035;		
Best Local Similarity	91.5%: Pred. No. 8.2e-181;		
Matches 906; Conservative	0; Mismatches 35; Indels 49; Gaps 3;		
QY	63	gaggactccagcgtgtcccaagtctgtgcaatctgcaactgtcgtccctcgacacctggaa	122
Db	18	GAGGACTCCAGCGTGGCCAGGTCGTGGCATCTTGACACTTACAGCCCTTGACACCTGGGAA	77
QY	123	gatggcggccgtggagccttcaaccctctctgtgtgttctgtgcaagcaccctgatca	182
Db	78	GATGCGCGCCCGTGGAGACTTACACCTTCTGTGTGTTGCTGCGCAGCCACCTTGATCCA	137
QY	183	agccaccttgcctccatgtcagttctatctctgtgcccacaaagtcatcaagaagaagct	242
Db	138	AGCCACCTTGAGTCCCATGTGACATGTTCTATCTCTGCGCCCAAAAGTCATCAAGAAGACT	197

OY	243	gacacaggagatcgtgaaggaccacaagaagcccaagaaactctctgcagagcgctggccgtcttaag	302
Db	198	GACACAGGAGCTGAAGAGCACCAACGCCACCAGCATCTCTGCAGCAGCGTCCGCTCTCAG	257
OY	303	tgcacatgcgggaaaaagccaagccgagag-attcccttgcttgggcaagcctgtgtgaacacct	361
Db	258	TGCCATGGGGGAAAAAGCCAGCGGAGGACATCCCTGTGCTGGGACACTGTTGAACACCGT	317
OY	362	cctgaagcacatcatctctgttgtgaagtatacaagctaacatcctccagcttcgcaagttaa	421
Db	318	CCTGAAGAACAATCAATCTGGCTGAAGGTATATCAACACTTAACATCTTCACAGCTGCAGGTAA	377
OY	422	gcccccggcccaaatacgaagagctctctgaagaatccccccgtggagacatggtgtcgtgaatt	481
Db	378	GCCCTCGGCCAATTAGCACGAGAGCTCTTAGTCAAGATCCCCCTGGACATGTTGGCTGGATT	437
OY	482	caacacgccccctgtgtcaagaaccatcgttgaagttccacaatacgaacttgaagcccaagcac	541
Db	438	CAACACGCCCCCTGGTCAAGAACCATGTTGAAGTTCACATGAGACTGAGGCCCAGGCAC	497
OY	542	catccgcattggaaccagatgacagtaggggccccaccgctgtgtccctcaatgaactgtgcac	601
Db	498	CATCGGCATGAGACACCAAGTGAGGCCCCACCGCCGGTGCCTCACTAGATGTTGGCAC	557
OY	602	caagcatlggaaacccctgcgatccaaactcgtgatagaactctccttcctcgtgtgaacgctt	661
Db	558	CAGCCATGGAGAGCTCGCCGATCCAACTGCTGATTAAGCTCTCTTCCTGTTGTAACGCTT	617
OY	662	agctaagaagttcataagaaacctctcagtgccalcctcctggcccaactagatgaanaaacagct	721
Db	618	AGCTAAGCAGGTCATGAACCTCTAGTGCATCCCTGCCAATCTAATGAAAACACAGCT	677
OY	722	gtgtccgctgatacgaaggtctccttaaitggaaatgatatcaagccctccgaagctgtgtaa	781
Db	678	GTGTCCCGGATCGAGGGCTCTTCAATGGCATGTATGAGAGACCTCTGCACACTGTGTA	737
OY	782	ggtgcacatltcccaccgaatcgaacggtctggaattgaaccttcgtalccctggcatcaa	841
Db	738	GG-----GTAGTGCTCTGCTGCTCTCTCCACATTTTTCC-----	772
OY	842	gggtgaacacattcagctctacccitgaggggccaagtlgttgtaactacaagggaaaagtgaac	901
Db	773	-----TTACTAAGGAGCTGGGCTTCACAGACCCGGGAAAAGTGAAC	809
OY	902	caagtggttacaataactctgcagactctcccttcgacaatgccaccctgggaacaatccgctt	961
Db	810	CANGTGTTCAATAACTCTGCAAGCTTCCCTGCATATGCCACCCCTGGAGACACATCCCGTT	869
OY	962	gaagctcatcgtgtagtcaaggaagctgtgtgaagaactcgaatlggtctgtgtctctccaga	1021
Db	870	CAGCCTCATCGTAGCATGACGAGGAGGTGTGAACCTGCAGTGGCTGTGCTCTTCAGAA	929
OY	1022	agaattcatgtctctgttggaactctgtgct	1051
Db	930	AGAAATTCATGTGCTCTGTGGACTCTGTGTGT	959
RESULT	7		
AX301906	AX301906	963 bp	DNA linear PAT 30-NOV-2001
LOCUS	Sequence 13 from Patent WO174851.		
DEFINITION	AX301906		
ACCESSION	AX301906.1 GI:17382956		
VERSION			
KEYWORDS	human.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Majumder,K., Spaderna,S.K., Taupier,R.J., Padigar,M.,		
	Burgess,C.E., Shimkets,R.A., Spytek,K.A., Liu,X., Patturajan,M. and		
	Gusev,V.Y.		



FEATURES	
source	Location/Qualifiers
repeat_region	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="20"
	/map=q11.1-11.23"
	/clone="RP5-1187J4"
	/clone_1lb="RPCI-5"
misc_feature	/note="157 copies 2 mer ga 61% conserved"
	/note=1824
mRNA	/note="Random repeat. weak data"
	join(3408..3471,5782..5867,7116..7295,8242..8295, 8799..8957,9841..9908,10729..10774,11885..11948, 12794..12870,15591..15761)
	/gene="d1187J4.1"
	/note="match: ESTs: Em:BE933124 Em:BE061388"
	/product="d1187J4.1.1 (novel protein similar to mouse von Ebner salivary gland protein, isoform 1.)"
	/evidence=not_experimental
gene	3408..15761
	/gene="d1187J4.1"
	join(<3408..3471,5782..5867,7116..7295,8242..8295, 8799..8957,9841..9908,10729..10774,11885..11948, 12794..12870,15591..15650)
	/gene="d1187J4.1"
	/note="Continues in Em:A1121901 as bA49G10.6"
	/codon_start=-1
CDS	/evidence=not_experimental
	/product="d1187J4.1.1 (novel protein similar to mouse von Ebner salivary gland protein, isoform 1.)"
	/protein_id="CAC34050.1"
	/db_xref="GI:13274680"
	/translation="LCPIVLEASFNGMVDILQIKVPISIDIRLEDILYPAKCDT IQLYIGAKLSDSGKYTKFNNSASLTMPFLDINPISLIYSQVKAANAATVSPEE EWLDLSVLPESAHRLKSITGLINEKADKLGSQTIVYLITDPPEFFIDGHAKVAQ LVLEVFPSEALRPLFTLGIEASEAQFYRKGDQLINLNMISSDRIOLMNCSIGMF QQDVKNITTEITHSIILPNONGKRSCGPVLVAFGEAAESSLTAKDALVTLPAST MKSSPVSO"
	3817..3915
misc_feature	/gene="d1187J4.1"
	/note="match: STS: Em:G10242"
	complement(join(3917..4006,31119..31326))
	/note="match: STS: Em:U75536"
	3917..4143
	/gene="d1187J4.1"
repeat_region	/note="match: STS: Em:G10242"
	4662..4691
	/note="15 copies 2 mer tt 86% conserved"
	join(5758..5867,7116..7295,8242..8295,8799..8957, 9841..9908,10729..10774,11885..11948,12794..12870, 15591..15761)
	/gene="d1187J4.1"
	/note="match: cDNAs: Em:U46068
misc_feature	match: ESTs: Em:A1924632 Em:AW363590 Em:AW384404 Em:AM449208 Em:AM167650 Em:AW381927 Em:AM167610 Em:AMS77643 Em:AM176323 Em:A1024451 Em:DA5555 Em:AM176329 Em:AM176327 Em:AM176325 Em:A1909965 Em:A1909964 Em:AA428580 Em:AW384405"
	/product="d1187J4.1.2 (novel protein similar to mouse von Ebner salivary gland protein, isoform 2.)"
	/evidence=not_experimental
	complement(6761..7253)
	/note="match: GSS: Em:AQ149447"
	join(7182..7295,8242..8295,8799..8957,9841..9908, 10729..10774,11885..11948,12794..12870,15591..15650)
misc_feature	/gene="d1187J4.1"
	/codon_start=-1
	/evidence=not_experimental
	/product="d1187J4.1.2 (novel protein similar to mouse von Ebner salivary gland protein, isoform 2.)"
	/protein_id="CAC15880.1"
	/db_xref="GI:11137741"
repeat_region	/translation="WPTLDNIPEFSITYSQDVKAANAATVSPEEFMWLDSVLPESA HRKSSIGLINERKADKLGSTQVIKTITDDPEEFFIDGHAKVAOLIVEFPPSSAAR PLFTLGIEASSEAOFTYRGDTILNLNNISSDRIOLMNCSIGMFOPVKIKNIIETH STLLPNONGKRSGVPSVLVAFGEAAESSLTAKDALVTLPASTMKRSSPVSO" 10970..11009
	/note="20 copies 2 mer ac 85% conserved"
	complement(11466..11766)
	/note="match: GSS: Em:AQ060606"
	complement(11631..11769)
	/note="match: GSS: Em:AQ427662"
misc_feature	11766..12192
	/gene="d1187J4.1"
	/note="match: GSS: Em:AQ403818"
	11768..12108
	/gene="d1187J4.1"
	/note="match: GSS: Em:AQ373301"
misc_feature	11775..12108
	/gene="d1187J4.1"
	/note="match: GSS: Em:AQ662246"
	14322..14833
	/note="match: GSS: Em:AQ62246"
	14836..14871
repeat_region	/note="256 copies 2 mer ca 68% conserved"
	14836..14871
	/note="18 copies 2 mer ac 88% conserved"
	complement(15031..15459)
	/note="match: GSS: Em:AQ694930"
	15736..15741
polya_signal	/gene="d1187J4.1"
	15754
	/gene="d1187J4.1"
	15761
	/gene="d1187J4.1"
	complement(17559..18000)
polya_site	/note="match: GSS: Em:AQ349609"
	complement(17617..18004)
	/note="match: GSS: Em:AQ350711"
	18109..18565
	/note="match: STS: Em:HS812D7T"
	complement(20661..21112)
misc_feature	/note="match: GSS: Em:AQ019396"
	complement(20698..21115)
	/note="match: GSS: Em:AQ178024"
	21105..21680
	/note="match: STS: Em:H5156A22T"
	complement(22577..22968)
misc_feature	/note="match: GSS: Em:AQ508829"
	22977..23573
	/note="match: GSS: Em:AQ281488"
	complement(37780..38128)
	/note="match: STS: Em:HSPFI3F11"
	complement(39112..39718)
misc_feature	/note="match: GSS: Em:AQC36647"
	complement(39266..39748)
	/note="match: GSS: Em:AQ827010"
	39747..40042
	/note="match: GSS: Em:AQ282137"
	39754..40132
misc_feature	/note="match: GSS: Em:AQ252720"
	39763..40037
	/note="match: GSS: Em:AQ029439
	match: STS: Em:G50113"
	complement(39950..40165)
	/note="match: GSS: Em:AL207848"
misc_feature	complement(39950..40334)
	/note="match: GSS: Em:AQ556676"
	39961..40217
	/note="match: GSS: Em:AQ143633"
	complement(39964..40169)
	/note="match: STS: Em:G02503"
misc_feature	complement(39991..40409)
	/note="match: GSS: Em:AQ596761"
	complement(40002..40172)
	/note="match: GSS: Em:AL185783"
	complement(40006..40317)







```

repeat_region /note="3 copies 60 mer 73% conserved"
               31921..32032
               /note="56 copies 2 mer ct 63% conserved"
repeat_region 33693..33762
               /note="MER47 repeat: matches 2251..2322 of consensus"
               33733..33785
               /note="MER47 repeat: matches 1..55 of consensus"
               join(355353..35585,36507..36678,40308..40452,41455..41562,
               42783..42935,45523..45604,46980..47043,47896..47936,
               48600..48787)
               /gene="BA49G10.1"
               /note="match: cDNAs: Em:U79414 Em:U79413 Em:AX061621
               match: ESTs: Em:A1654622"
               /product="BA49G10.1 (similar to bovine salivary protein
               BSP30)"
gene /evidence=not_experimental
     35535..48787
     /gene="BA49G10.1"
     join(36522..36678,40308..40452,41455..41562,42783..42935,
     45523..45604,46980..47043,47896..47936)
     /gene="BA49G10.1"
     /note="match: proteins: Tr:P79125 Sw:P07743"
     /codon_start=1
     /evidence=not_experimental
     /product="BA49G10.1 (similar to bovine salivary protein
     BSP30)"
     /protein_id="CAC03546.1"
     /db_xref="GI:9801234"
     /translation="MLQLMKLVLCGLVTGSESLDNLGNLDSNVWDEKLEPEVLEGL
     ETVNNTLKGILEKLVLDGLQSSAKMOIAKQAEAKLLNNVSKLLPNTDFGL
     KINSLLIDVKAEPIDDGKGLNLSPPVANTYVAGPIIGGIINLKASLDLTAATIEF
     DPQHPVAVVGECASDPTSLSLDKHSDIINNFVSNVITLKSYSLLQKICP
     LIRFIHSLDVNDVNOQVNDPNQHKQLOTLI"
repeat_region 38155..38751
     /note="LIPa8 repeat: matches 4514..5125 of consensus"
repeat_region 38750..39785
     /note="LIPa8 repeat: matches 5128..6163 of consensus"
polyA_signal 48769..48774
     /gene="BA49G10.1"
     48787
     /gene="BA49G10.1"
     52540..53450
     /note="Cpg island"
     /evidence=not_experimental
misc_feature complement(53831..54131)
     /note="match: GSS: Em:A0897745"
misc_feature 53831..54141
     /note="match: STS: Em:G31731"
     53842..53969
     /note="match: GSS: Em:A0429347"
     complement(53844..54141)
     /note="match: GSS: Em:AF101963"
     53858..54022
     /note="match: GSS: Em:AQ392891"
     complement(53908..54141)
     /note="match: GSS: Em:AQ415287"
     53958..54139
     /note="match: GSS: Em:AQ378081"
     53970..54139
     /note="match: GSS: Em:AQ527291"
     53972..54142
     /note="match: GSS: Em:AQ429347"
     54000..54141
     /note="match: GSS: Em:AQ392510"
     complement(54077..54139)
     /note="match: GSS: Em:AQ533658"
     54030..54138
     /note="match: GSS: Em:AC000935"
     54041..54149
     /note="match: GSS: Em:AQ392891"
     54767..55668
     /gene="BA49G10.2"
     54767..55668
     CDS

```

```

Query Match          9.6%; Score 161; DB 9; Length 161593;
Best Local Similarity 100.0%; Pred. No. 1,6e-27;
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 78 ccacgtctgcatcctgcatctgctgcccctctgacacctgggaagatgcccgcgctg 137
      |||||||
DB 153404 ccacagctcggcatcctgcatctgctgcccctctgacacctgggaagatgcccgcgctg 153463
      |||||||
QY 138 gaccttcacccctctctgtgtgttctgtgacgacacccctgtatccacccctcaagtc 197
      |||||||
DB 153464 GACCTTCACCCCTCTCTGTGTGTGCTGTGACACACCTTGATCAAGCCACCTCAGTCC 153523
      |||||||
QY 198 cactgcagcttcacccctctgctgcccacaaagtcataaagaa 238
      |||||||
DB 153524 CACTGCAGTTCACCTCTCTGCGCCCAAAAGTCATCAAGAA 153564

RESULT 10
AC094188/c
LOCUS AC094188/c
DEFINITION Rattus norvegicus clone CH230-2K16, *** SEQUENCING IN PROGRESS ***,
49 unordered pieces.
ACCESSION AC094188.2 GI:17940901
VERSION AC094188.2
KEYWORDS HTG: HTGS PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 89536)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Alt-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbarta,D.,
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Bubay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyte,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Demn,A.L., Ding,Y., Dinh,H.H., Douthaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Barnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homs,F., Howard,S., Huber,D., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,D., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kuresh,A., Landry,N., Deal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martinale,A.,
Martinez,E., Massey,E., Mawhiney,E., Mcleod,M.P., Mendor,M.,
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkeno,S.,
Oguni,M., Okunou,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Plimus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolle,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshatali,N.,
Sisson,I., Sodegren,E., Sonalke,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabot,P., Tamerisa,K., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,

```

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,  
Wellington, S., Williams, G., Williamson, A., Mieczyk, R., Wooden, S.,  
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G. and Gibbs, R.  
Direct Submission  
Unpublished  
2 (bases 1 to 89536)

Worley, R.C.  
Submitted (17-SEP-2001)

Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Dec 20, 2001 this sequence version replaced gi:15636537.

## COMMENT

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GAEB

Center clone name: CH230-2K16

----- Summary Statistics -----  
Assembly program: Phrap; version 0.990329first call to  
findphraplist

Consensus quality: 73144 bases at least Q40  
Consensus quality: 80017 bases at least Q30  
Consensus quality: 85368 bases at least Q20  
Estimated insert size: 69357; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-gel estimation  
Quality coverage: 0.8x in Q20 bases; sum-of-contigs estimation

----- NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
NOTE: This is a 'working draft' sequence. It currently  
consists of 49 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 4713: contig of 4713 bp in length  
4714 4813: gap of unknown length  
4814 7442: contig of 2629 bp in length  
7443 7543: gap of unknown length  
7543 8970: contig of 1428 bp in length  
8971 9070: gap of unknown length  
9071 11618: contig of 2548 bp in length  
11619 11718: gap of unknown length  
11719 14064: contig of 2346 bp in length  
14065 14164: gap of unknown length  
14165 16598: contig of 2434 bp in length  
16599 16598: gap of unknown length  
16599 18558: contig of 1860 bp in length  
18559 18558: gap of unknown length  
18559 21636: contig of 2978 bp in length  
21637 21736: gap of unknown length  
21737 23091: contig of 1355 bp in length  
23092 23191: gap of unknown length  
23192 24363: contig of 1172 bp in length  
24364 24463: gap of unknown length  
24464 26616: contig of 2153 bp in length  
26617 26716: gap of unknown length  
26717 29328: contig of 2612 bp in length  
29329 29429: gap of unknown length  
29430 30861: contig of 1433 bp in length  
30862 30961: gap of unknown length  
30962 32666: contig of 1705 bp in length  
32667 32766: gap of unknown length  
32767 34751: contig of 1985 bp in length  
34752 34851: gap of unknown length  
34852 36722: contig of 1871 bp in length  
36723 36822: gap of unknown length

## FEATURES

## source

36823 37990: contig of 1168 bp in length  
37991 38091: gap of unknown length  
38091 39852: contig of 1761 bp in length  
39852 39952: gap of unknown length  
39952 41483: contig of 1531 bp in length  
41483 41582: gap of unknown length  
41582 42970: contig of 1388 bp in length  
42970 43071: gap of unknown length  
43071 44071: contig of 1647 bp in length  
44071 44718: gap of unknown length  
44718 44818: gap of unknown length  
44818 46732: contig of 1914 bp in length  
46732 46831: gap of unknown length  
46831 48721: contig of 1890 bp in length  
48721 48822: gap of unknown length  
48822 50484: contig of 1663 bp in length  
50484 50585: gap of unknown length  
50585 52334: contig of 1749 bp in length  
52334 52434: gap of unknown length  
52434 53718: contig of 1285 bp in length  
53718 53818: gap of unknown length  
53818 55081: contig of 1263 bp in length  
55081 55181: gap of unknown length  
55181 56437: contig of 1256 bp in length  
56437 56538: gap of unknown length  
56538 58395: contig of 1858 bp in length  
58395 58495: gap of unknown length  
58495 60176: contig of 1681 bp in length  
60176 60277: gap of unknown length  
60277 62186: contig of 1910 bp in length  
62186 62286: gap of unknown length  
62286 64498: contig of 2212 bp in length  
64498 64598: gap of unknown length  
64598 65899: contig of 1391 bp in length  
65899 66089: gap of unknown length  
66089 67493: contig of 1404 bp in length  
67493 67593: gap of unknown length  
67593 67594: gap of unknown length  
67594 69641: contig of 2048 bp in length  
69641 69741: gap of unknown length  
69741 70862: contig of 1121 bp in length  
70862 70962: gap of unknown length  
70962 72389: contig of 1427 bp in length  
72389 72490: gap of unknown length  
72490 74083: contig of 1593 bp in length  
74083 74182: gap of unknown length  
74182 75219: contig of 1036 bp in length  
75219 75319: gap of unknown length  
75319 76803: contig of 1491 bp in length  
76803 76909: gap of unknown length  
76909 78708: contig of 1799 bp in length  
78708 78808: gap of unknown length  
78808 79849: contig of 1041 bp in length  
79849 79949: gap of unknown length  
79949 81538: contig of 1589 bp in length  
81538 81638: gap of unknown length  
81638 82748: contig of 1110 bp in length  
82748 82848: gap of unknown length  
82848 84058: contig of 1210 bp in length  
84058 84158: gap of unknown length  
84158 85695: contig of 1537 bp in length  
85695 85795: gap of unknown length  
85795 86875: contig of 1080 bp in length  
86875 86975: gap of unknown length  
86975 88063: contig of 1088 bp in length  
88063 88163: gap of unknown length  
88163 89536: contig of 1373 bp in length.

Location/Qualifiers

1..89536

/organism="Rattus norvegicus"

/db\_xref="taxon:10116"

/clone="CH230-2K16"

## BASE COUNT

21918 a 20067 c 20357 g 22306 t 4888 others



Matches	119:	Conservative	0:	Mismatches	64:	Indels	0:	Gaps	0:
QY	1081	gcacgcgagctgacaaatgaagaagctgcagatgaagctggatgctaccagatcggtgaaga							1140
Db	38807	GGATGTGTGTGATGCTTTCACAGCGACCAACAAATTGGGGGCCACACCAGATGTTGACAA							38748
QY	1141	tccctaactaagaacactcccgagctttttatatagaccaaagggcgaatgagcccaac							1200
Db	38747	TCTTCACTACAGCACCCCCCATATTTTGTAAATGAGGGCGGTGCCAGCGACGCCANA							38688
QY	1201	tgatcgatgcgtgaagatggtttccctccagtgaaagccctccgcctttgttaccctgggaca							1260
Db	38687	GGTGTGTACTCGAAGATGTTCCCCACACCAATACAGATGTCCGACCCCTTCTTCGCGCA							38628
QY	1261	tccg							
Db	38627	TTTG							

RESULT	12
AC107771	
LOCUS	69591 bp DNA
DEFINITION	linear
ACCESSION	AC107771
VERSION	AC107771.1
KEYWORDS	HTG; HTGS_PHASED.
SOURCE	house mouse.
ORGANISM	Mus musculus

REFERENCE  
Mammalia, Eutrachia, Chrotalia, Craniata, Vertebrata; Euteleostomi;  
Atherinomorpha, Euthrinacina, Rodentia; Sciuromorpha; Muridae; Mus.  
REFERENCE  
AUTHORS  
1 (bases 1 to 68591)  
Birren, B., Linton, L., Nusbaum, C., and Lander, E.  
TITLE  
Mus musculus, clone RP23-106C21  
JOURNAL  
Unpublished  
2 (bases 1 to 68591)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.

**TITLE** Direct Submission  
**JOURNAL** Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** All repeats were identified using RepeatMasker.

Smit, A.F.A. & Green, P. (1996-1997)  
<http://www.1754.com>

----- Genome Center  
[lcp://lcp.genome.washington.edu/RM/RepeatMasker.html](http://lcp.genome.washington.edu/RM/RepeatMasker.html)

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WITR

Web site: <http://www-seq.wi.mit.edu>

----- Project Information -----  
 contact: sequence\_submissions@genome.wi.mit.edu

Center project name: L204333

```
-----
Center clone name: 106_C_21
```

\* NOTE: This record contains 82 individual

```
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
714 813: contig of 713 bp in length
814 1570: gap of 100 bp
1571 1670: contig of 757 bp in length
1671 2427: contig of 757 bp in length
2428 2527: gap of 100 bp
2528 3286: contig of 759 bp in length
3287 3386: gap of 100 bp
3387 4108: contig of 722 bp in length
4109 4208: gap of 100 bp
4209 4945: contig of 737 bp in length
4946 5045: gap of 100 bp
5046 5771: contig of 726 bp in length
5772 5871: gap of 100 bp
5872 6606: contig of 735 bp in length
6607 6706: gap of 100 bp
6707 7430: contig of 724 bp in length
7431 7530: gap of 100 bp
7531 8279: contig of 749 bp in length
8280 8379: gap of 100 bp
8380 9134: contig of 755 bp in length
9135 9234: gap of 100 bp
9235 9954: contig of 720 bp in length
9955 10054: gap of 100 bp
10055 10804: contig of 750 bp in length
10805 10904: gap of 100 bp
10905 11644: contig of 740 bp in length
11645 11744: gap of 100 bp
11745 12458: contig of 714 bp in length
12459 12558: gap of 100 bp
12559 13290: contig of 732 bp in length
13291 13390: gap of 100 bp
13391 14120: contig of 730 bp in length
14121 14220: gap of 100 bp
14221 14959: contig of 739 bp in length
14960 15059: gap of 100 bp
15060 15814: contig of 755 bp in length
15815 15914: gap of 100 bp
15915 16663: contig of 749 bp in length
16664 16763: gap of 100 bp
16764 17506: contig of 743 bp in length
17507 17606: gap of 100 bp
17607 18359: contig of 753 bp in length
18360 18459: gap of 100 bp
18460 19211: contig of 752 bp in length
19212 19311: gap of 100 bp
19312 20036: contig of 725 bp in length
20037 20136: gap of 100 bp
20137 20870: contig of 734 bp in length
20871 20970: gap of 100 bp
20971 21703: contig of 733 bp in length
21704 21803: gap of 100 bp
21804 22550: contig of 747 bp in length
22551 22650: gap of 100 bp
22651 23388: contig of 738 bp in length
23389 23488: gap of 100 bp
23489 24227: contig of 739 bp in length
24228 24327: gap of 100 bp
24328 25079: contig of 752 bp in length
25080 25179: gap of 100 bp
25180 25924: contig of 745 bp in length
25925 26024: gap of 100 bp
26025 26757: contig of 733 bp in length

26758 26857: gap of 100 bp
26858 27608: contig of 751 bp in length
27609 27708: gap of 100 bp
27709 28456: contig of 748 bp in length
28457 28556: gap of 100 bp
28557 29305: contig of 749 bp in length
29306 29405: gap of 100 bp
29406 30140: contig of 735 bp in length
30141 30240: gap of 100 bp
30241 30938: contig of 698 bp in length
30939 31038: gap of 100 bp
31039 31747: contig of 709 bp in length
31748 31847: gap of 100 bp
31848 32577: contig of 730 bp in length
32578 32677: gap of 100 bp
32678 33439: contig of 762 bp in length
33440 33539: gap of 100 bp
33540 34268: contig of 729 bp in length
34269 34368: gap of 100 bp
34369 35100: contig of 732 bp in length
35101 35200: gap of 100 bp
35201 35945: contig of 745 bp in length
35946 36045: gap of 100 bp
36046 36795: contig of 750 bp in length
36796 36895: gap of 100 bp
36896 37624: contig of 729 bp in length
37625 37724: gap of 100 bp
37725 38456: contig of 732 bp in length
38457 38556: gap of 100 bp
38557 39274: contig of 718 bp in length
39275 39374: gap of 100 bp
39375 40119: contig of 745 bp in length
40120 40219: gap of 100 bp
40220 40944: contig of 725 bp in length
40945 41044: gap of 100 bp
41045 41791: contig of 747 bp in length
41792 41891: gap of 100 bp
41892 42644: contig of 753 bp in length
42645 42744: gap of 100 bp
42745 43468: contig of 724 bp in length
43469 43568: gap of 100 bp
43569 44318: contig of 750 bp in length
44319 44418: gap of 100 bp
44419 45168: contig of 750 bp in length
45169 45268: gap of 100 bp
45269 46013: contig of 745 bp in length
46014 46113: gap of 100 bp
46114 46827: contig of 714 bp in length
46828 46927: gap of 100 bp
46928 47632: contig of 705 bp in length
47633 47732: gap of 100 bp
47733 48460: contig of 728 bp in length
48461 48560: gap of 100 bp
48561 49286: contig of 726 bp in length
49287 49386: gap of 100 bp
49387 50127: contig of 741 bp in length
50128 50227: gap of 100 bp
50228 50981: contig of 754 bp in length
50982 51081: gap of 100 bp
51082 51821: contig of 740 bp in length
51822 51921: gap of 100 bp
51922 52658: contig of 737 bp in length
52659 52758: gap of 100 bp
52759 53510: contig of 752 bp in length
53511 53610: gap of 100 bp
53611 54354: contig of 744 bp in length
54355 54454: gap of 100 bp
54455 55195: contig of 741 bp in length
55196 55295: gap of 100 bp
55296 56024: contig of 729 bp in length
56025 56124: gap of 100 bp
56125 56874: contig of 750 bp in length
56875 56974: gap of 100 bp
```

		56975	57694: contig of 720 bp in length
	*	57695	gap of 100 bp
Query Match		4.6%;	Score 77.6; DB 2; Length 68591;
Best Local Similarity		63.8%;	Pred. No. 1e-07; Mismatches 74; Indels 2; Gaps 1
Matches 134; Conservative		0;	
QY 188	cctcagtcacactgtagtctcatctctcgtgcgccaaagtcatacaagaanaagctgacac	247	
DB 51568	ccccccccccccccccccatcttgatcatttaaacagcattctctctgcagacgtgacc	51627	
QY 248	aggagctgaagaccacacaacgacacacagatctctgacagagctgcgcgtctcaagtcca	307	
DB 51628	agcgactgaagagacattgatgtccactgcctatctctccagagatgtgcacgtctcaagcca	51687	
QY 308	tgcgggaaaaagccagccggaagatccctctgtctgtgcacgcgtgtgaaacacgcttctgaa	367	
DB 51688	tgcgaagatnaagctc--gcagatccccaactactgacacagcttcgtagcaccacgcttctgag	51745	
QY 368	gcacatctctgctcgaagatcatcaagc	397	
DB 51746	atcactcatatgctgagtgaggccgcagcagc	51775	

RESULT	13
LOCUS	AC094188
DEFINITION	AC094188 89536 bp DNA linear HTG 20-DEC-2001
ACCESSION	AC094188
VERSION	AC094188.2 GI:17940901
KEYWORDS	HTG; HTGS PHASE1.
SOURCE	Norway rat.
ORGANISM	Rattus norvegicus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
	Rattus.
	1 (bases 1 to 89536)
REFERENCE	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
AUTHORS	Alshrocks,S.L., Amaralunga,H.C., Aze,J.R., Banks,T., Barbarella,J.,
	Benton,J., Bimege,K., Blankenburg,K., Bonnin,D., Bouck,J.,
	Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Bubay,C.,
	Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,
	Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
	Chen,Z., Chowdry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
	Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
	Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
	Dem,A.L., Ding,Y., Dinh,H., Douthwaite,K.J., Diaper,H.,
	Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
	Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
	Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
	Garra,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
	Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A.,
	Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Hollway,C.,
	Hollins,B., Homs,J.F., Howard,S., Huber,J., Huylk,S., Hume,J.,
	Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
	Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,T.,
	Kovar,C., Kirtlovic,J., Kursh,A., Landry,N., Leal,B., Lewis,L.C.,
	Lewis,L., Li,J., Li,Z., Lichtenarge,O., Lieu,C., Liu,J., Liu,W.,
	Loiseged,H., Locado,R.J., Lu,X., Luchter,A., Lucier,R., Luna,R.,
	Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
	Martinez,E., Massey,E., Mawhinney,E., Mcleod,M.P., Meador,M.,
	Mel,G., Metzker,M., Miner,G., Minter,Z., Mitchell,T., Mohabot,K.,
	Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
	Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkenwo,S.,
	Ogah,M., Okwono,G., Orangunye,N., Oyiedo,K., Pace,A., Payton,B.,
	Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
	Roules,M., Ren,Y., Rives,M., Rojas,A., Rojebokan,I., Rolfe,M.,
	Sisson,I., Sodergren,E., Sonshke,T., Sparks,A., Stanely,H.,
	Stone,H., Sutton,A., Svatek,A., Taber,P., Tameris,A., Tameris,K.,
	Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
	Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Wali, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,  
 Wellington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S.,  
 Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstock, G., and Gibbs, R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 89536)  
 Morley, K.C.  
 Direct Submission  
 Submitted (17-SEP-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Dec 20, 2001 this sequence version replaced gi:15636537.  
 ----- Genome Center

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
-----  
Project Information  
Center project name: GAEB  
Center clone name: CH230-2K16  
-----  
Summary Statistics  
Assembly Program: Phrap, version 0.990329First call to  
indphraplist

Consensus quality: 73144 bases at least Q40  
Consensus quality: 80017 bases at least Q30  
Consensus quality: 85368 bases at least Q20  
Estimated insert size: 69357; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-1p estimation  
Quality coverage: 0.8x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length (see [http://www.bncs.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.bncs.bcm.tmc.edu/docs/genbank_draft_data.html)).  
NOTE: This is a 'working draft' sequence. It currently consists of 49 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

*	1	4713:	contig of 4713	bp in length
*	4774	4813:	gap of unknown	length
*	4814	7442:	contig of 2629	bp in length
*	7443	7542:	gap of unknown	length
*	7543	8970:	contig of 1428	bp in length
*	8971	9070:	gap of unknown	length
*	9071	11618:	contig of 2548	bp in length
*	11619	11718:	gap of unknown	length
*	11719	14064:	contig of 2346	bp in length
*	14065	14166:	gap of unknown	length
*	14165	16598:	contig of 2434	bp in length
*	16599	16599:	gap of unknown	length
*	16599	18588:	contig of 1860	bp in length
*	18589	18658:	gap of unknown	length
*	18659	21636:	contig of 2978	bp in length
*	21637	21736:	gap of unknown	length
*	21737	23091:	contig of 1355	bp in length
*	23092	23191:	gap of unknown	length
*	23192	24363:	contig of 1172	bp in length
*	24364	24616:	gap of unknown	length
*	24616	26616:	contig of 2153	bp in length
*	26617	29338:	gap of unknown	length
*	29339	29428:	contig of 2612	bp in length
*	29429	30861:	contig of 1433	bp in length
*	30862	30961:	gap of unknown	length
*	30962	32666:	contig of 1705	bp in length
*	32667	32766:	gap of unknown	length
*	32767	34761:	contig of 1985	bp in length
*	34752	34851:	gap of unknown	length
*	34852	36722:	contig of 1871	bp in length
*	36723	36822:	gap of unknown	length

Query Match	3.5% ; Score 59.4 ; DB 2 ; length 89536 ;
Best Local Similarity	77.4% ; Pred. No.0.0021 ;
Matches 72 ;	Conservative 0 ; Mismatches 21 ; Indels 0 ; Gaps 0 ;
Qy 633	cataagctctctctctctgtagaacgccttagcctaagcagcagtcataacctaagcttagcgca 692
Db 65004	CAAAAGCCTCTCTCTCTGCTGTCACACTCTCTGGCAAAAGACGTTATGCAATCTCTGATGCGCA 65063
Oy 693	tccttgcccaatctagtagaanaaccagctggt 725
Db 65064	GCCCTGCCCAAAATAGTGAATAAGCCACCTGAGT 65096
RESULT 14	
LOCUS	AF429315 125020 bp DNA linear PRI 18-JAN-2002
DEFINITION	Homo sapiens juncctophillin 3 (JPH3) gene, partial cds.
ACCESSION	AF429315
VERSION	AF429315.1 GI:17646244
KEYWORDS	
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 125020) Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S., Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A., Potter,N.T., Ross,C.A. and Margolis,R.L.
TITLE	A repeat expansion in the gene encoding juncctophillin-3 is associated with Huntington disease-like 2
JOURNAL	Nat. Genet. 29 (4), 377-378 (2001)
MEDLINE	2 (bases 1 to 125020)
PUBMED	11583737
REFERENCE	11594876
AUTHORS	Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
TITLE	Direct Submission
JOURNAL	Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES	Location/Qualifiers
source	1..125020
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="16"
	/map="16q24.3; between D16S520 and WI-12410"
	/note="isolated from a patient with Huntington's disease-like 2 (HDL2)"
	complement(35581..35746)
	/rpt_type=tandem
	/rpt_unit=ctg
repeat_region	complement(<36507..>36887)
mrna	/gene="JPH3"
	/product="juncctophillin 3"
gene	complement(<36507..>36887)
	/gene="JPH3"
	/note="JPH3"
CDS	complement(<36507..36887)
	/gene="JPH3"
	/note="component of the junctional complex between plasma membrane and endoplasmic reticulum"
	/codon_start=1
	/product="juncctophillin 3"
	/protein_id="AA140941.1"
	/db_xref="GI:17646245"
	/translation="MSSGGRFNPDDGSGYCGGMEGKAHGHCVTGKRGGEYTGSMWSHGFEVLVTPMISGNTYQGTWADCRHGIGLESKGVWYGEWTHGFRGVEGACGNGAEKGTWNSGLDDGYGTEYSNG"





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 06:28:35 ; Search time 3813.96 Seconds

(without alignments)  
1728.350 Million cell updates/sec

Title: US-09-700-770-4

Perfect score: 315  
Sequence: 1 taaacactgactcaattt.....tggtattactactatgc 315

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pal:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sy:\*  
12: gb\_un:\*  
13: gb\_vl:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

C	1	315	100.0	118068	9	AL356798	AL356798 Human DNA
C	2	21	6.7	3705	14	NPXKA	M16821 Autographa
C	3	21	6.7	133694	6	A48542	A48542 Sequence 1
C	4	21	6.7	133894	14	L22858	L22858 Autographa
C	5	21	6.7	142289	9	AC078955	AC078955 Homo sapi
C	6	21	6.7	163283	9	AL355606	AL355606 Human DNA
C	7	20	6.3	57287	9	AC008283	AC008283 Homo sapi
C	8	20	6.3	81580	8	AP000736	AP000736 Arabidops
C	9	20	6.3	87966	2	AC105559	AC105559 Rattus no
C	10	20	6.3	98017	8	AC027033	AC027033 Arabidops
C	11	20	6.3	109981	2	AC096828	AC096828 Rattus no
C	12	20	6.3	111159	9	AC008271	AC008271 Homo sapi
C	13	20	6.3	114927	2	AC105629	AC105629 Rattus no
C	14	20	6.3	11548	2	AC094962	AC094962 Rattus no
C	15	20	6.3	117757	9	HSU100419	AL121912 Human DNA
C	16	20	6.3	139619	2	AL627164	AL627164 Danio rer
C	17	20	6.3	146950	2	AC095100	AC095100 Rattus no
C	18	20	6.3	152695	2	AC099576	AC099576 Mus muscu
C	19	20	6.3	161394	2	AC010759	AC010759 Homo sapi
C	20	20	6.3	211681	2	AC083915	AC083915 Mus muscu
C	21	19	6.0	512	11	AF220278	AF220278 Oncorhyn
C	22	19	6.0	912	8	AB011794	AB011794 Citrus si
C	23	19	6.0	1042	10	S78689	S78689 LGI-1=large
C	24	19	6.0	5313	8	SCYNR054C	Z71669 S.cerevisia
C	25	19	6.0	7310	8	SCYNR055C	Z71670 S.cerevisia
C	26	19	6.0	36148	3	U39740	U39740 Caenorhadi
C	27	19	6.0	82316	9	AL354726	AL354726 Human DNA
C	28	19	6.0	83552	9	AL157759	AL157759 Human DNA
C	29	19	6.0	101410	8	ATAC009540	AC009540 Arabidops
C	30	19	6.0	125245	2	AC097581	AC097581 Rattus no
C	31	19	6.0	138228	2	AC108885	AC108885 Rattus no
C	32	19	6.0	140935	9	AC025173	AC025173 Homo sapi
C	33	19	6.0	144493	9	AC097721	AC097721 Homo sapi
C	34	19	6.0	145496	2	AC080176	AC080176 Homo sapi
C	35	19	6.0	150001	9	AC006063	AC006063 Homo sapi
C	36	19	6.0	152631	2	AC048373	AC048373 Homo sapi
C	37	19	6.0	154307	2	AC108150	AC108150 Homo sapi
C	38	19	6.0	157332	2	AC093250	AC093250 Homo sapi
C	39	19	6.0	161059	2	AC098610	AC098610 Rattus no
C	40	19	6.0	161824	9	AC096721	AC096721 Homo sapi
C	41	19	6.0	175448	2	AC009662	AC009662 Homo sapi
C	42	19	6.0	177027	9	AC097514	AC097514 Homo sapi
C	43	19	6.0	177429	9	AP002378	AP002378 Homo sapi
C	44	19	6.0	178310	2	AC023884	AC023884 Homo sapi
C	45	19	6.0	180185	2	AC010763	AC010763 Homo sapi

## ALIGNMENTS

RESULT 1  
LOCUS AL356798/c  
DEFINITION Human DNA sequence from clone RP11-467B11 on chromosome 9, complete sequence.  
ACCESSION AL356798  
VERSION AL356798.18 GI:15795413  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 118068)  
AUTHORS Babbage/A.  
TITLE Direct Submission  
JOURNAL Submitted (26-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
On Sep 27, 2001 this sequence version replaced gi:14455883.  
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: EMBL, SW, SWISSPROT, TrEMBL, WPI, MORNREP, Information on the MORNREP database can be found at

http://www.sanger.ac.uk/projects/C\_elegans/wormrep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr9  
Rp11-467B11 is from the library RPC1-11.2 constructed by the group of Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone Rp11-467B11. It may be shorter because we sequence overlapping sections only once, except for a short overlap. The true right end of clone Rp11-467B11 is at 118068 in this sequence. The true left end of clone Rp11-208P1 is at 29578 in this sequence. The true right end of clone Rp11-29B11 is at 2000 in this sequence.

#### FEATURES

source  
1..118068  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="9"  
/clone="Rp11-467B11"  
/clone\_id="RPC1-11.2"  
BASE COUNT 33053 a 23625 c 22510 g 38880 t  
ORIGIN

#### Query Match

Best Local Similarity 100.0%; Score 315; DB 9; Length 118068;  
Matches 315; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taacactgactcagatttaagaataactttgagaataagaacaatgaatcagtt 60  
Db 96845 TAAACAGCATTCAGATTTTAAGAAATTAAGTTGAGAAATGAAATGAAATCAGTT 96786  
QY 61 tctccaccacttaagatatactcttagagatactacagccctcccttagggagatacaaa 120  
Db 96785 TCTCCACCACCTTAAGTATCTCTTAAAGATCTACAGCCCTCTTAGGGGACATCAAA 96726  
QY 121 gtcagtgctgttgccttggtagtccaccattatcaagtagatgatacaaat 180  
Db 96725 GTCAGTGTGCTGCTTGTGAGTCCACCTTATATTCAGTAGATGATGACATAAAT 96666  
QY 181 ttgaataatagatgacacataaactgagtttagaagaacatcagtagaagaataa 240  
Db 96665 TTGAAATATAGATTGTACACAAATAACTGAGATTATGGAACATCAGTAGAAGAAATA 96606  
QY 241 caacattccatccctttagagagatacttacttgaactcaggaataattgtcattgt 300  
Db 96605 CAACATTCATCCCTTTAGAGAGATCTTACTTGCAACTCAGGATAATTTGTCATGTGT 96546  
QY 301 attatctactatgc 315  
Db 96545 ATTATCTACTTATGC 96531

RESULT 2  
NPHKEA/c 3705 bp DNA linear VRL 07-FEB-1994  
LOCUS NPHKEA  
DEFINITION Autographa californica nuclear polyhedrosis virus HindIII K region

ACCESSION M16821  
VERSION M16821.1 GI:332445  
KEYWORDS 35 kDa protein; 94 kDa protein.  
SOURCE Autographa californica nuclear polyhedrosis virus (strain L-1) DNA, passed in Spodoptera frugiperda (IPB-SF-21 cells).  
ORGANISM Autographa californica nucleopolyhedrovirus  
VIRUSES; dsDNA viruses, no RNA stage; Baculoviridae;  
Nucleopolyhedrovirus.

REFERENCE Friesen, P.D. and Miller, L.K.  
1 (bases 1 to 3705)  
Divergent transcription of early 35- and 94-kilodalton protein genes encoded by the HindIII K genome fragment of the baculovirus Autographa californica nuclear polyhedrosis virus

JOURNAL Virol. 61, 2264-2272 (1987)  
MEDLINE 87226411  
COMMENT Displaced messenger-active RNAs are transcribed from both 35- and 94-kD protein genes early (2 hours) after infection. Late in infection, abundant RNAs are transcribed from promoters located at least 2.5 kb upstream from the gene encoding the 35 kD protein. These transcripts completely overlap both the 35 and 94 kD polypeptide genes, but apparently lack protein-coding potential, suggesting that the transcripts may play a role in suppressing early viral gene expression.

#### FEATURES

source  
location/Qualifiers  
1..3705  
/organism="Autographa californica nucleopolyhedrovirus"  
/db\_xref="taxon:46015"  
complement(39..2684)  
/note="early 94 kb mRNA (alt.)"  
complement(39..2503)  
/note="early 94 kb mRNA (alt.)"  
complement(39..2569)  
/note="early 94 kb mRNA (alt.)"  
complement(79..2490)  
/note="early 94 kb protein"  
/codon\_start=1  
/protein\_id="AAA46702.1"  
/db\_xref="GI:332446"

#### mrna

translation="MINTYVATDSDLSSTNSDYFNKNALQTLFQFONEIENISSCDK  
VLIHMSVACQKGTGDKVSKYLRNNGSGDTFRSEFIKWIKHKNINLIDKILLYV  
TDGQISKEANVCNKLNEKPFSEFVYFAINNTEQIDLSVASAVNNSSDKTYRN  
DEWVWVNIKTEPNYDILITTEENFISKDELLSFREFEINSMPDANVINEVYKRL  
RORLSEIKOTNNSMNFDOIKNNKEPVPKSPFEYKTYNADVLNPDITDSTIS  
AIYLNHNKSYAFDVMKNLHYNKLSVARETDDVYANDDEAYDYSNENRFPDC  
ILANDSGPATILLHYNLEPITQSLTKRFSRDEPLVMSQNEIKNSIIECYNEEL  
KOLIOHGRISPRSRPPTGAIVNEPQFDEVDNVLACTEYDAKVFANGLWYLYL  
KHINDATIDNVADYFRKRYIYRINTECMIGISNLAPEPLIKVLPALWYSEIS  
TLERHONOHGKELKQFAHFAEDMIQIOWCQYTDVNEAVKRAYCLKRIIMFR  
MSVDAVEMIANRAFECKDEKFIINLNADALQDLKFKVNHNGYVDEHVLNDSINA  
ERYLYFVHIIEDFKYSYVDNMPRPAFVEEGKTPYDSILLKQLOSYPHNGQITTEK  
CSRDFENRILSLHRLYIECVKSLNKYPTLEFYQNYNNOCHYFNKRIAPPEIIONL  
AAVHNEYANKIVNLPVEEPIYRANNYRITRIONERVSPLDAEETIDKILKISEQV  
NICRK"

#### cds

2667..3607  
/note="early 35 kb mRNA (3' end +/- 3 bp)"  
2701..3600  
/note="early 35 kb protein"  
/codon\_start=1  
/protein\_id="AAA46703.1"  
/db\_xref="GI:332447"  
/translation="KCVIFPVEIDVSTIIRDCQVDKQRELYVINKINNTQITKPVL  
MMPNIGSIRSVTRKNNNLDRIKSVDEDFQLEBDYSDOMGFDSDIKYKDEHS  
VSCQNGSVLSKSKFAKILKSHDYDDEKKSIAVEYCLPKLVDERNDYVAVVCLKGEF  
NGSNQVLSFEYFNIGNKVLYPFAHEINQGLYEYDVAVYVDSVQDGRQFEPFVSLI  
LPSSFKNSEKVLVYNENSKNSMITYKALEFTTSSNGKSEKIMKIFCGFIYDKRSK  
VLVYKLNHTSALNKVILNTIK"

BASE COUNT 1080 a 686 c 731 g 1208 t  
ORIGIN 122 bp upstream of HindIII site; 85-88 mu on the viral map.

Query Match  
Best Local Similarity 100.0%; Score 21; DB 14; Length 3705;  
Matches 21; Conservative 2.5;

```

Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 21 aagaataactttgagaat 41
Db 464 AAGAAATAACTTTTGAGAAAT 444

RESULT 3
AA8542/c 133894 bp DNA linear PAT 07-MAR-1997
LOCUS Sequence 1 from Patent WO9601320.
ACCESSION A48542
VERSION A48542.1 GI:2302312
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 133894)
AUTHORS Bishop,D., Possee,R. and Ayres,M.
TITLE AUTOGRAPHA CALIFORNICA COMPLETE GENOME SEQUENCE
JOURNAL PATENT: WO 9601320-A 1 18-JAN-1996;
NATURAL ENVIRONMENT RES (GB)
COMMENT Other publication AU 2897295 960125.
FEATURES
source
1. 133894
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 39195 a 27151 c 27347 g 40201 t
ORIGIN

Query Match 6.7%; Score 21; DB 6; Length 133894;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 21 aagaataactttgagaat 41
Db 114255 AAGAAATAACTTTTGAGAAAT 114235

RESULT 4
L22858/c 133894 bp DNA circular VRL 29-MAR-2001
LOCUS Autographa californica nucleopolyhedrovirus clone C6, complete
DEFINITION genome.
ACCESSION L22858
VERSION L22858
KEYWORDS L22858.1 GI:510708
SOURCE Autographa californica nucleopolyhedrovirus.
ORGANISM Autographa californica nucleopolyhedrovirus.
REFERENCE 1 (bases 1 to 133894)
AUTHORS Ayres,M.D., Howard,S.C., Kuzio,J., Lopez-Ferber,M. and Possee,R.D.
TITLE The complete DNA sequence of Autographa californica nuclear
polyhedrosis virus
JOURNAL Virology 202 (2), 586-605 (1994)
MEDLINE 94303173
PUBMED 8030224
REFERENCE 2 (bases 1 to 133894)
AUTHORS Kuzio,J.
TITLE Direct Submission
JOURNAL Submitted (08-MAR-1999) NCBI, Bethesda, MD 20894, USA
FEATURES
source
1. 133894
/organism="Autographa californica nucleopolyhedrovirus"
/db_xref="taxon:46015"
/clone="C6"
503. 1009
/gene="Ac-ptp"
/gene="Acorf-1"
503. 1009
/gene="Ac-ptp"
CDS

```

```

/note="PTP: 19288 Da primary translation product"
/codon_start=1
/product="protein tyrosine phosphatase"
/protein_id="AA66631.1"
/db_xref="GI:559070"
/translation="MEPARMHWYLOCGGVIKDSNLICRPTLPRLPELFAVYVSEEDVMT
AEQIKONPSIGAILDLNFTSKSYDGVGLYKKIQVGGTLPPESIVQEFIDT
VKEEFKCPGMLVGVCHTNGINRRTGYMCRYLMHTLGIAPOGALDRFEKAGHKIERQ
NYVDLLT"
complement(1041..2027)
/gene="Ac-bro"
/note="Ac-bro"
complement(1041..2027)
/gene="Ac-bro"
/note="Ac-bro"
/codon_start=1
/product="BRO: 37769 Da primary translation product"
/protein_id="AAA6632.1"
/db_xref="GI:559071"
/translation="MARVKIGFERGEGEDPNLRVYLERRDQVREYAKVANSIKRTVC
DKAIRVHDNKKSLFEQTIQNGPSTSVYKRGDPLQHPVLTITSGVQLINKS
KLPIAIEIQLLEEVITPOVLCTGKTDPAIKOREESQVLTKLIATTEHTNALQAV
VAQKTEELVKRQEFLETVAIKDKQIEAKDLQVKTVMIDLMNMYGFGQETMOKDEIM
OKKDAQVLDLVAKVVDLSDRAVOYPADRKRHVLCTVRDGTTFATITGQKTVENOKH
KRNIIVANIVENIRPNPTVDNNATDRLQAKRSRSIYLVWKKRRNNLKIG"
complement(2084..2245)
/gene="Ac-ctx"
/note="Ac-ctx"
complement(2084..2245)
/gene="Ac-ctx"
/note="Ac-ctx"
/codon_start=1
/product="CTX: 5590 Da primary translation product"
/protein_id="AAA6633.1"
/db_xref="GI:559072"
/translation="MQIKTVLLAFAMFALNAQHVLAACAEAGVAVHNDECCGAGCS
PIFNVCPLQ"
complement(2295..2750)
/gene="Acorf-4"
complement(2295..2750)
/gene="Acorf-4"
/note="Acorf-4"
complement(2295..2750)
/gene="Acorf-4"
/note="Acorf-4"
/codon_start=1
/product="Acorf-4 peptide"
/protein_id="AAA6634.1"
/db_xref="GI:559073"
/translation="MKLYTKNASLKYALRLTREKENITIPFDLTLRLRLDIDMKIK
SEDVQRFNRTNNDLISACMQINVRTYFMFNATIDMRKQNPCTYFRICQCHLEADVPS
PDDHSVYRFLCVACGTPVLDHPDVFQHTGEGVNLLEVOGVNAGGEL"
2779..3108
/gene="Acorf-5"
2779..3108
/gene="Acorf-5"
/note="Acorf-5"
complement(12435 Da primary translation product"
/codon_start=1
/product="Acorf-5 peptide"
/protein_id="AAA6635.1"
/db_xref="GI:559074"
/translation="MRTSRINNAPVVASQHDYRDQIKRELNSLRVNHVICTRSCT
SFDCKNFLRSDMDTPVVTITPKRTADYKITEYVGDVATIKRSPNPLVESGLVREAA
KYGECIV"
3089..3721
/gene="Ac-lef2"
/note="Ac-lef2"
complement(1041..2027)
/gene="Ac-lef2"
/note="Ac-lef2"
/codon_start=1
/product="late expression factor 2"
/protein_id="AAA6636.1"
/db_xref="GI:559075"
/translation="MANASYNWSPLIRASCLDKRATYVLDIDPDDFIDKLTLLPTVTFY
NGVLVTKISGLRLMLTAPPTINIKSNFRRSKRNKCMKECVEGKKNVDMLNK

```

```

gene
/INMPICIKILNDKENNENRGMYKKRFFILNOCYLANVSCACENRCLIKALTHFYN
complement(3759..4364)
/gene="Ac-ORF603"
/note="Ac-ORF-7"
complement(3759..4364)
/gene="Ac-ORF603"
/note="ORF603 peptide; 23612 Da primary translation
product"
/codon_start=1
/protein_id="AAA66637.1"
/db_xref="GI:559076"
/translation="MAVLENNKQLADNSIEKGELEFLNGSYNTLSEYNPVLLKNG
VLELEAAVYAGNIIKYTDDEPFIDYINLIKATSEELPENSIVYVNRKTMRSCTIH
PIKDIYIYDNKFTLYDRYIYGYDNVNVNVEEKEKEKEEEDKASSLCENKII
LSQINCESEFENFKYLYSDYNVAFSLIDNTNIVAFGLYR"
4520..5257
/gene="Ac-PH"
/note="Ac-ORF-8"
4520..5257
/gene="Ac-PH"
/note="polyhedrin; 28642 Da primary translation product"
/codon_start=1
/product="major occlusion body protein"
/db_xref="GI:559077"
/translation="WPDYSYRPTIGRTVYVYDNKYKMLGAVIKAKRKHFAHEHEIE
ATLDPDLYVAEDPFLGPGKQKLTLEKIRANYPPTMKLVGMKGEKREYRETFWPF
MEDSEPIYNDQEDVFLVYVMRPTRPKRCFKLAQALHCDPPYVHDYIRVSEBM
VGSNNERYISLAKGGGCGPILNLSHSEYTNSEQIDVAVINENTFKPLVYIGTDSAE
EILLESIVFVKYEFAPDAPLFTGPAT"
complement(5287..6918)
/gene="Ac-ORF1629"
/note="Ac-ORF-9; Ac-61k; Ac-vp78"
complement(5287..6918)
/gene="Ac-ORF1629"
/note="61k; vp78; required for virus replication; 60713 Da
primary translation product"
/codon_start=1
/product="viral capsid associated protein"
/protein_id="AAA66639.1"
/db_xref="GI:559078"
/translation="MTNRRYSEVSYLNNRNKIDAHOFEEFVDFAAOIKNNIYD
NIVNLDVYLNLIKLANDVFNKAYVYDSDSVSHYVAVYMKRQICVGRPSIRQ
SLVNTAIYERLNLIGVNDSEITMLIADYDLYSVNIEFLPQALPSRRSVQOP
AAPAVPTIVREQTKPEOILRAPPPSPVNIIPAPPPSPKSELPPAPMPVPEP
SPPPPPPPPPPPAPPPKAVKTRKRVKTAPESTVVEPTVLPKTFEPKPPSA
VRKPAQKRQSEIIPKSSSTNLADVLITNRRVAMAKSSSEATSNDEGDDDDN
RPNKANTPDVYQALFNVTSSQLYTNDSEDRNTKAHNLINDVEPLTQNTQINDK
ARLLQDLASFVLSNPDLSPAIGSEKQPLFETNNLTKSTIEDLIFRKYDAENH
LIPALYHPDXYFENLKYQOLSVNQRTSSA"
6917..7735
/gene="Ac-PK-1"
/note="Ac-ORF-10"
6917..7735
/gene="Ac-PK-1"
/note="PK1; 31978 Da primary translation product"
/codon_start=1
/product="protein kinase"
/protein_id="AAA66640.1"
/db_xref="GI:559079"
/translation="MATNATLQTLVQFENCKNVKTRKIIINGRFGKISILSHKPS
KLYLQKTSIAHNENADEIKVHOLMSDHPNFIKITYNHGSINNVOYVMYDIDPDEF
LOIKGELSYOLVSNIRIOLCEALNDLKHNFHNDIKLENIYEPALDRVYVCDGIC
KHENSLSYDGTLEVSEPKIRHTTMHVSFPMYANGVTLKLLTGGRHPEKSEDEML
DLSMKRRQOVNDIVLKHVRVNAADVYLCITRINIDCRLLNKKIILKHEFLS"
7747..7864
/standard_name="h1a"
/note="2 copies of 30 bp imperfect palindromic sequence"
/function="enhancer; replication origin"
/rpl_type="dispersed
complement(7899..8921)
repeat_region
gene
```

```

CDS
/gene="Ac-ORF-11"
complement(7899..8921)
/gene="Ac-ORF-11"
/note="40093 Da primary translation product"
/codon_start=1
/product="Ac-ORF-11 peptide"
/protein_id="AAA66641.1"
/db_xref="GI:559080"
/translation="MSIAAKLIIYVYAKNEVHDVYGSYHHYRIYQVLSBSYVNG
MSCIERVYTMARLKSQSCFEDPAVYKIDGDSIKLSHPSFTSETMGIDNVREYLE
QIDAVYVSVYVONQOITSUNNEREISODMIDLCQIILGREFYMRKGLRLIANY
FNRNDVGVMMYKFCVVIYVHIMRSVPAELSEAVKFIILRSKSDYDLRLHV
DESYNCEPVYAEYGRFCGIGKHEFSKRLSCWHILFOYLRGKTQDEEKSPCYAVIK
DFQRCQKDYVKNLKDVEDLILHAHSMDSKDKNSIMDLLCYMDCBEIDVDFYIIFESFL
NNKHQ"
8958..9611
/gene="Ac-ORF-12"
8958..9611
/gene="Ac-ORF-12"
/note="25412 Da primary translation product"
/codon_start=1
/product="Ac-ORF-12 peptide"
/protein_id="AAA66642.1"

Query Match      6.7%; Score 21; DB 14; Length 133894;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      21 aagaataactttgagaat 41
Db 114255 AAGAATAACTTTGAGAAAT 114235

RESULT      5
AC078955 LOCUS      AC078955 Homo sapiens 142289 bp DNA linear PRI 30-MAR-2001
DEFINITION      Homo sapiens 12q BAC RP11-543H12 (Roswell Park Cancer Institute
ACCESSION      AC078955 Human BAC library) complete sequence.
VERSION      AC078955.24 GI:13489125
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 142289)
REFERENCE      Muzny D.M., Adams C., Adio-Oduola B., Ali-osman F.R., Allen C.,
AUTHORS      Alsbrooks S.L., Amaratunge H.C., Are J.R., Banks T., Barbara J.,
Benton J., Blmage K., Blankenburg C., Bonnin D., Bouck J.,
Bowle S., Brileva M., Brown E., Brown M., Bryant N.P., Buhay C.,
Burche P., Burkett C., Burrell K.L., Byrd N.C., Carton T.F.,
Carter M., Cavazos S.R., Chacko J., Chavez D., Chen G., Chen R.,
Chen Z., Chiu D., Chowdhry I., Christopoulos C., Cleveland C.D.,
Cox C., Coyle M.D., Dathorne S.R., David R., Davila M.L., Davis C.,
Day-Carroll L., Dederich D.A., Delaney K.R., Delgado O.,
Deem A.L., Ding Y., Dinh H.H., Douthwaite K.J., Draper H.,
Dugan-Rocha S., Durbin K.J., Earnhart C., Edgar D., Edwards C.C.,
Elhaj C., Emerling S., Escotto M., Falls T., Ferraguto D.,
Flagg N., Ford J., Foster P., Frantz P., Gabisi A., Gao J.,
Garcia A., Garner T., Garza N., Gill R., Gorrell J.H., Guevara W.,
Gunnarone P., Hale S., Hamilton K., Han J., Harris C., Harris K.,
Hart M., Haylak P., Hawes A., Hernandez J., Hernandez O.,
Hodgson A., Hogue M., Hollway C., Hollins B., Homsi F.,
Howard S., Huber J., Hulyk S., Hume J., Ioshikhes I., Jackson L.E.,
Jacobson B., Jia Y., Johnson R., Jolivet S., Joudah S.,
Karlsson E., Kelly S., Khan U., King L., Korah J., Kovar C.,
Kratovic J., Kunesli A., Landry N., Leal B., Lee E., Lewis L.C.,
Lewis L., Li J., Li Z., Licharge O., Lien C., Liu J., Liu W.,
Louisgeed H., Lozado R.J., Lu X., Luchter A., Luchter R.,
Ma J., Maheshwari M., Mapua P., Marondel I., Matlin R.,
Martindale A., Martinez E., Massey E., Mawhney E., McLeod M.P.,
Meador M., Mel G., Merscher S., Metzker M., Miller A., Miner G.,
Miner Z., Mitchell T., Mohabbat K., Montgomery K.T., Morgan M.,
```

TITLE	Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, N., Nguyen, N., Nguyen, N., Nickerson, E., Nockentwo, S., Oguh, M., Okwona, G., Oragunye, N., Oyedero, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shm, C., Shoohtlati, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, E., Stone, H., Sutton, A., Svatek, A., Tabori, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalona, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorilla, S., Zucherlapati, R. and Gibbs, R.
REFERENCE	Direct Submission
REFERENCE	Unpublished
TITLE	2 (bases 1 to 142289)
AUTHORS	Worley, K. C.
JOURNAL	Direct Submission Submitted (13-AUG-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (bases 1 to 142289)
AUTHORS	Worley, K. C.
TITLE	Direct Submission
JOURNAL	Submitted (30-MAR-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT	On Mar 30, 2001 this sequence version replaced g1:13470106. INFORMATION: <a href="http://www.hgsc.bcm.tmc.edu/">http://www.hgsc.bcm.tmc.edu/</a> or email <a href="mailto:gc-help@bcm.tmc.edu">gc-help@bcm.tmc.edu</a>

**CLONE LENGTH:** This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

**ANNOTATION OF FEATURES:**  
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-14) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

Summary Statistics	
Contig length:	142289
Phrap values in estimate:	141529
Average error rate (BCM-Phrap estimate):	1.31902e-05

Fraction of Phrap values less than 40 :	0.00705862
Number of consensus changing edits:	41
Number of N's in consensus :	0

Position	Original+Context	Edited+Context
51	aacattatag(n)attttttg	aacattatg(t)atttttg
52	atttttttg(n)ttgttttg	attttttg(t)ttgttttg
67	ttttttttg(n)ttgttttt	ttttttttg(t)ttgttttt
3797	taatttaagt(n)aacacaaa	taatttaagt(t)aacacaaa
17427	ccctccctc(n)acatggcac	ccctccctc(t)aacatggcc
41447	ggatccaggaa(n)anncggtg	ggatccaggat(t)accggtg
41448	ttgtctaggan(a)nncggtgt	ttgtctaggat(t)accggtgt
41449	acctaggana(n)ncggtgtg	acctaggat(t)ccggtgtg
41450	gtccagganaa(n)cggtgtgtg	gtccaggat(a)cggtgtgtg
41631	ltaaggaggaa(n)accaggagcc	ltaaggaggat(g)accaggagcc
43314	ctctgctagc(n)acatggagc	ctctgctagt(t)acatggagc
68569	ggatctagact(n)ngttaaana	ggatctagat(a)gattaaana
68571	gtagagact(n)ltaanaaat	gtagagagt(n)ltaanaaat
68576	actnngttta(n)aaatgtg	actnngttat(g)aaatgtg
68587	aaaatgtgc(n)catnnttgc	aaaatgtgc(a)catnnttgc
68591	lgtgcncn(n)lntnccatg	lgtgcncat(n)lntnccatg
68593	ttgtgcact(n)lntccatggaa	ttgtgcactat(a)lntccatggaa
68595	ngcnatnnt(n)ccatggatn	ngcnatnnt(n)ccatggatn
68605	nccatggat(n)ntltgagcc	nccatggat(a)ntltgagcc
68742	ccatggaatn(n)tatgcagcca	ccatggaat(c)tatgcagcca
74888	tyggatcttga(n)ctaatggatc	tyggatcttga(t)ctaatggatc
74891	ggagctccgt(n)ltaaaaaa	ggagctccgt(t)ltaaaaaa
75765	actccgttgc(n)aaaaaaaaa	actccgttgc(a)aaaaaaaaa
77345	lgaataaaaa(n)agatataaaa	lgaataaaaa(g)agatataaaa
77346	lgaagcaac(n)nnntnnac	lgaagcaac(c)nnntnnac
77347	ggagcaaac(n)nnntnnac	ggagcaaac(a)nnntnnac
77348	cagcaaacm(n)ntnnaaat	cagcaaac(a)ntnnaaat
77349	agcaaacnm(n)ntnnaaatg	agcaaac(a)ntnnaaatg
77351	gaacnmnn(n)ltnnaaaatga	gaacnmnn(a)ltnnaaaatga
77352	gaacnmnn(n)naaatgaaat	gaacnmnn(t)naaatgaaat
77353	gaacnmnn(n)naatgaaat	gaacnmnn(t)naatgaaat
77389	gcaagtgtg(n)naatgncc	gcaagtgtg(c)naatgncc
77390	ccagttgtg(n)naatgncc	ccagttgtg(c)naatgncc
77391	agcgttgn(n)atgtgncc	agcgttgn(t)atgtgncc
77396	agcgttgn(n)ccctaaact	agcgttgn(a)ccctaaact
92451	aaactaaatt(n)naatlgagc	aaactaaatt(t)naatlgagc
109360	acctaaatn(n)naatlgaga	acctaaat(a)naatlgaga
119939	agtgaataa(c)ccgcacaag	agtgaataa(a)ccgcacaag
120033	lcttgagtt(n)cttgaagata	lcttgagtt(t)cttgaagata
	actctgagc(n)ltagcagctc	actctgagc(a)ltagcagctc

----- Distribution of Quality < 40 Bases

#	Phrap Value Range
1000	
900	
800	
700	
600	
500	
400	
300	
200	
100	
0	

```

Version: 1.01  qxf0.
FEATURES
  Location/Qualifiers
    source          1. .142289
    /organism="Homo sapiens"

```

```
1.01 qxf.
Location/Qualifiers
1.14289
/organism="Homo sapiens"
```

/db\_xref="taxon:9606"  
/chromosome="12g"  
/clone="RP11-543H12"  
1..2000  
/note="Overlaps bases 42818..44817 of clone AC004550"  
/function="Overlaps with adjacent clone AC004550"  
53..80  
/rpt\_family="TTTTG)n"  
complement(81..362)  
/rpt\_family="AluY"  
complement(597..967)  
/rpt\_family="L2"  
1019..1811  
/rpt\_family="L1PA16"  
3063..3347  
/rpt\_family="AluSg"

Query Match 6.7%; Score 21; DB 9; Length 142289;  
Best Local Similarity 100.0%; Pred. No. 2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 gaatagacaatgaatca 57  
|||||  
Db 119773 GAAATGAGACAAATGAAATCA 119793

RESULT 6  
AL355606/c 163283 bp DNA linear PRI 09-MAR-2001  
LOCUS Human DNA sequence from clone RP11-9017 on chromosome X, complete  
DEFINITION  
ACCESSION AL355606  
VERSION AL355606  
KEYWORDS HTG: GI:13273693  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Mankayota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
TITLE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 163283)

JOURNML  
COMMENT Direct Submission  
Submitted (08-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
On Mar 11, 2001 this sequence version replaced gi:12742783.  
During sequence assembly data is compared from overlapping clones.  
Where difference are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em: EMBL; Sw:  
SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP  
database can be found at  
http://www.sanger.ac.uk/projects/c-elegans/wormpep

was generated from part of bacterial clone configs of human  
chromosome X, constructed by the Sanger Centre Chromosome X Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/ChX

RP11-9017 is from the library RP11-11.1 constructed by the group of  
Pietier de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBAC3.6

FEATURES  
This sequence is the entire insert of clone RP11-9017.  
Location/Qualifiers

source  
1..163283  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/clone="RP11-9017"  
/clone\_11b="RP11-11.1"  
1004..1071  
/note="2 copies 34 mer 89% conserved"  
2756..3638  
/note="CpG island"  
/evidence=not\_experimental  
3737..3816  
/note="L2 repeat: matches 2100..2179 of consensus"  
3861..3983  
/note="L2 repeat: matches 2049..2180 of consensus"  
3975..4059  
/note="MIR repeat: matches 15..102 of consensus"  
4060..4360  
/note="AluSx repeat: matches 1..303 of consensus"  
4361..4534  
/note="MIR repeat: matches 102..262 of consensus"  
4559..4694  
/note="L2 repeat: matches 2208..2355 of consensus"  
5091..5388  
/note="AluSx repeat: matches 1..299 of consensus"  
6254..6612  
/note="MER67C repeat: matches 47..424 of consensus"  
6719..7019  
/note="L2 repeat: matches 2217..2500 of consensus"  
7049..7515  
/note="MER31B repeat: matches 1..485 of consensus"  
7755..8058  
/note="AluIo repeat: matches 1..309 of consensus"  
8421..9284  
/note="L2 repeat: matches 902..1813 of consensus"  
9346..9630  
/note="AluSg repeat: matches 1..293 of consensus"  
9794..9919  
/note="L2 repeat: matches 2626..2750 of consensus"  
10518..10971  
/note="L2 repeat: matches 1832..2298 of consensus"  
11443..11738  
/note="AluY repeat: matches 1..297 of consensus"  
12805..13029  
/note="AluIo repeat: matches 86..311 of consensus"  
13148..13910  
/note="L1PA16 repeat: matches 5383..6157 of consensus"  
13907..14705  
/note="L1PA7 repeat: matches 5347..6143 of consensus"  
14704..14957  
/note="L1P repeat: matches 4046..4299 of consensus"  
14958..15248  
/note="L1P repeat: matches 4603..4893 of consensus"  
15240..15575  
/note="L1P repeat: matches 4285..4621 of consensus"  
15572..15938  
/note="L1P3 repeat: matches 1..4059 of consensus"  
19940..20174  
/note="L1P repeat: matches 5155..5398 of consensus"  
20181..20208  
/note="tRNA-Gly-GCC repeat: matches 1..28 of consensus"  
21901..22068  
/note="MT1J repeat: matches 1..165 of consensus"  
22727..22780  
/note="L2 repeat: matches 2699..2750 of consensus"  
23407..23594  
/note="MIR repeat: matches 48..256 of consensus"  
23601..23767  
/note="MER5A repeat: matches 4..153 of consensus"  
24307..24871  
/note="MT1E repeat: matches 1..553 of consensus"  
25135..25364  
/note="MT1J repeat: matches 109..319 of consensus"

repeat\_region 25433. .25505  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 25872. .25911  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 25913. .26066  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 26170. .26464  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 26822. .27035  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 28226. .28682  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 28683. .29029  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 29055. .30453  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 30454. .31023  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 31038. .31154  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 31364. .31448  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 31455. .34275  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 34276. .34571  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 34572. .35031  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 35039. .35092  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 35095. .35842  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 35853. .39029  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 39031. .39643  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 39646. .40249  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 40250. .40552  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 40553. .42188  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 42508. .42799  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 42800. .43029  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 43201. .43306  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 43778. .44031  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 44186. .44361  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 44468. .44799  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 44930. .45100  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 45981. .46564  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 46605. .46734  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 47030. .47196  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 47575. .48669  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 48671. .50424  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 50940. .51248  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 51249. .51339  
/note="L1M4A repeat: matches 319. .402 of consensus"  
repeat\_region 52657. .53183  
/note="L1M4A repeat: matches 319. .402 of consensus"

repeat\_region 53242. .53938  
/note="L1M4A repeat: matches 5325. .5870 of consensus"  
repeat\_region 53939. .54218  
/note="L1M4A repeat: matches 5325. .5870 of consensus"  
repeat\_region 54828. .55070  
/note="L1M4A repeat: matches 5325. .5870 of consensus"  
repeat\_region 56554. .57365  
/note="L1M4A repeat: matches 5325. .5870 of consensus"  
Query Match 6.7%; Score 21; DB 9; Length 163283;  
Best Local Similarity 100.0%; Pred. No. 2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 112 acatacaagtcagttggtt 132  
Db 75453 ACATACAAAGTCAGTTGTTT 75433  
RESULT 7  
AC008283/c 57287 bp DNA linear PRI 21-DEC-1999  
LOCUS Homo sapiens BAC clone RP11-575G1 from 7q11.2-q22, complete  
DEFINITION sequence.  
ACCESSION AC008283  
VERSION AC008283.3 GI:5836161  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 57287)  
AUTHORS Sulston, J.E. and Waterston, R.  
TITLE Toward a complete human genome sequence  
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
MEDLINE 99063792  
REFERENCE 2 (bases 1 to 57287)  
AUTHORS Cordum, H., Stoneking, T., Nguyen, C., and Harris, A.  
TITLE The sequence of Homo sapiens BAC clone RP11-575G1  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 57287)  
AUTHORS Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (31-JUN-1999) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
REFERENCE 4 (bases 1 to 57287)  
AUTHORS Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (08-SEP-1999) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
REFERENCE 5 (bases 1 to 57287)  
AUTHORS Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (28-SEP-1999) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
REFERENCE 6 (bases 1 to 57287)  
AUTHORS Waterston, R.  
TITLE Direct Submission  
JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Sep 8, 1999 this sequence version replaced gi:5732142.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc  
Contact: sapiens@atston.wustl.edu  
----- Summary Statistics  
Center project name: H\_NH0575G01  
-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nih.gov/DIR/OTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenno, M., Catanese, J.J., and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://paccpac.med.buffalo.edu>)

#### VECTOR:

BAC3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP4-56004, 200 bp overlap: the clone sequenced to the right is RP5-117919, 200 bp overlap. The actual start of this clone is at base position 80345 of RP4-56004; actual end is at base position 76835 of RP5-117919.

#### Location/Qualifiers

#### FEATURES

##### source

```
1..57287
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q11.2-q22"
/clone="RP11-575G1"
/clone_1lb="RPC1-11"
1..257
/rpt_family="MER1_type"
285..309
/rpt_family="AT_rich"
2332..2753
/rpt_family="L2"
3005..3078
/rpt_family="MIR"
3223..3517
/rpt_family="Alu"
3838..3903
/rpt_family="AT_rich"
complement(<4103..>50924)
/genes="WUGSC:H_NH0575601.1"
complement(join(<4103..4261,26454..26553,30330..30380,
32481..32550,50795..>50924))
/genes="WUGSC:H_NH0575601.1"
/note="match to JH0565 (PID:g87003); H_NH0575601.1"
/codon_start=1
/product="match to calcium channel alpha-2b chain
precursor"
/protein_id="AA19234.1"
/db_xref="gi:6624059"
translation="TIVLNLMTSALDEVEKRNREDDPSLIMVFGSATGIARYPA
SPWVNSRTPKIDLYVRRRPWYIQGAAPKDLIIVDSGSGITLILIRISSE
MLETISDDDFVNAVSPNSMDVSCFQHLVQAVNRKKVLLDAVNITANGITDYKKG
FSFAFEQLN"
```

```
misc_feature
4111..4261
/note="match to EST N85205 (NID:g1260830)"
4212..4256
/note="match to EST A1434840 (NID:g4298992) t113a04.x1"
5073..5153
/rpt_family="MIR"
5343..5402
/rpt_family="L2"
5457..5581
/rpt_family="MIR"
5671..6718
/rpt_family="L2"
6719..7034
/rpt_family="Alu"
7035..7558
/rpt_family="L2"
8711..9067
/rpt_family="L2"
10066..10094
/rpt_family="AT_rich"
10095..10209
/rpt_family="Alu"
10532..10555
/rpt_family="AT_rich"
10602..10653
/rpt_family="AT_rich"
11251..11256
/note="match to EST A121707 (NID:g1679348) zn80a01.r1"
11733..11927
/rpt_family="MIR"
13680..13997
/rpt_family="L2"
15639..15782
/rpt_family="MER1_type"
16263..16553
/rpt_family="Retroviral"
16364..16723
/note="match to EST A1223011 (NID:g3805214) g952e11.x1"
16810..16953
/rpt_family="MIR"
18853..18988
/rpt_family="MIR"
19017..19295
/rpt_family="Alu"
19457..19498
/rpt_family="AT_rich"
20431..20674
/rpt_family="Retroviral"
21005..21318
/rpt_family="Retroviral"
21406..21941
/rpt_family="L1"
21942..21974
/rpt_family="TA)n"
22036..22057
/rpt_family="(A)n"
22294..23118
/rpt_family="Retroviral"
23152..23172
/rpt_family="AT_rich"
23205..23294
/rpt_family="MER1_type"
24112..24354
/rpt_family="Retroviral"
25734..26165
/rpt_family="L1"
26167..26270
/rpt_family="(TA)n"
26450..26553
/note="match to EST A1434840 (NID:g4298992) t113a04.x1"
26454..26539
/note="match to EST N85205 (NID:g1260830)"
26648..26675
repeat_region
```



```

repeat_region      27603..28308      /rpt_family="Ar_rich"

Query Match
Best Local Similarity 100.0%, Pred. No. 7.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      5      cactgactcagatttaaga 24
          |||
Db      15492 CACTGACTCAGATTTAAGA 15473

RESULT      8
LOCUS      AP000736      81580 bp      DNA      linear      PLN 27-DEC-2000
DEFINITION  Arabidopsis thaliana genomic DNA, chromosome 3, TAC clone:K17E7.
ACCESSION  AP000736 BA000014
VERSION    AP000736.3 GI:8843906
KEYWORDS
SOURCE
ORGANISM   Arabidopsis thaliana (strain:columbia) DNA, clone_lib:Mitsui TAC
           clone:K17E7.
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
           Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE  1 (sites)
AUTHORS    Kaneko,T., Katoh,T., Sato,S., Nakamura,A., Asamizu,E. and Tabata,S.
TITLE      Structural analysis of Arabidopsis thaliana chromosome 3. II.
           Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC
           and BAC clones
JOURNAL    DNA Res. 7 (3), 217-221 (2000)
MEDLINE    20363099
REFERENCE  2 (bases 1 to 81580)
AUTHORS    Kaneko,T., Kato,T., Sato,S., Nakamura,Y., Asamizu,E. and Tabata,S.
TITLE      Direct Submission
JOURNAL    Submitted (18-NOV-1999) Yasukazu Nakamura, Kazusa DNA Research
           Institute, Department of Plant Gene Research, 1532-3, Yana,
           Kisarazu, Chiba 292-0812, Japan (E-mail:ynakam@kazusa.or.jp.
           Tel:81-438-52-3935, Fax:81-438-52-3934)
           On Jun 29, 2000 this sequence version replaced gi:6451951.
           Address for correspondence: kaos@kazusa.or.jp
           For the latest information on annotation of this clone, please see
           http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=K17E7
           Genes with similarity to proteins in the databases are described in
           'product' or 'note' qualifiers. Genes that have no significant
           protein similarity are described as 'unknown protein'.
           The software programs used to predict genes include: Graal
           (Informatics Group, Oak Ridge National Laboratory,
           http://compbio.ornl.gov/Graal-1.3/),
           GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
           NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
           Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
           SplicePredictor (Volker Brendel, Stanford University,
           http://gremli1.1.zo1.lastate.edu/cgi-bin/sp.cgi).
           Genes encoding tRNAs are predicted by tRNAscan-SE
           (Sean Eddy, Washington University School of Medicine, St. Louis,
           http://genome.wustl.edu/eddy/tRNAscan-SE).
           This sequence may not be the entire insert of this clone. It may be
           shorter because we remove overlaps between neighboring submissions.
           The 5' clone is T26612 and the 3' clone is T20F20.
FEATURES
source      Location/Qualifiers
           1..81580
           /organism="Arabidopsis thaliana"
           /strain="Columbia"
           /db_xref="taxon:3702"
           /chromosome="3"
           /clone="K17E7"
           /clone_lib="Mitsui TAC"
           955..5085
           /note="contains similarity to retroelement pol
           yprotein
           gene_id:K17E7.1"
           /codon_start=1

CDS
/pseudo
/evidence-not experimental
join(6774..6898,6993..7357,7439..7782)
/note="gene_id:K17E7.2"
/codon_start=1
/pseudo
/evidence-not experimental
/product="transposon protein-like"
join(8339..8392,9478..8714,8838..9266)
/note="contains similarity to En/Spm transposon protein
gene_id:K17E7.3"
/codon_start=1
/evidence-not experimental
/protein_id="BAB02991.1"
/db_xref="GI:13477078"
/translation="MHKHLAGQKTYMQVYLQEMEBDLKRRVSGEYPMKTHRYDGFV
DQKSHVGEAYMKMLEERKSEIDKDGPOISGNSQSHRMLSIDKNEIFKCTILD
GKGNFGLSEKLVETLNKRRTESVSSSTVQLOLOLAKSEDAEQAKCDEHRQ
SOSRIASLEKLVLFKKKDPDELDAVLSDSPTDEEVTLPTTHHTSNRHSQPTNHRH
RHRRQOIRHHTLSLYFSYEV"
complement(join(9810..10123,10217..10337))
/note="emb|CAB88116.1
gene_id:K17E7.4
similar to unknown protein"
/codon_start=1
/evidence-not experimental
/protein_id="BAB02992.1"
/db_xref="GI:13477079"
/translation="MLHSDGLFAFLGESNGCKEMKLVSDAKDMOSQVSESDIV
NLVVEDSEIDEDCSIMGPKRKSGSIFPKPKRETTTIIQENLQFAPILVPMWLQ
TLRLRLSDIDPSLNKKVAKKSPISDELRTSVSRKATV"
join(112471..12477,12881..13057,13190..13277,13381..13486)
/note="gene_id:K17E7.5
unknown protein"
/codon_start=1
/evidence-not experimental
/protein_id="BAB02993.1"
/db_xref="GI:13477080"
/translation="MIYAMNAISFNAGIGFGRSNTPTPSESTTRATYTHRSVSAI
EVEALIALALQARSGLGFSFIPSSANCAADLLAHALCAFPSNSVPRIVEKQKSTIG
VKRSKQSGTHHPSSSSSANFIL"
join(14293..14724,14767..15851,15918..16290,16333..17136,
17203..18789)
/note="gene_id:K17E7.6"
/codon_start=1
/pseudo
/evidence-not experimental
/product="retroelement pol
yprotein-like"
complement(20744..29799)
/note="contains similarity to TNP2-like transposon
protein
gene_id:K17E7.7"
/codon_start=1
/pseudo
/evidence-not experimental
join(25276..25546,26136..26343,26806..27063,27478..27585,
27647..27820,28233..28440,29579..29743)
/note="gene_id:K17E7.8
unknown protein"
/codon_start=1
/evidence-not experimental
/protein_id="BAB02994.1"
/db_xref="GI:13477081"
/translation="MTYQFPNCLANRGKXSSONLRNSFLNMPPRDLFVGEDEV
CRDLFVGEDEVCRDLFVGKFEVCGENSSNLSFIRBNSSANLKITSSAKKFRV
EITYSSAKKREVEIYSSANLRFVGENSSANLSFIGNNLSANLSFIRBNSSANLSS
FLAIVSQCEGNIRKRVCDGIAWSCSGEITYSKKRVAFREIYSKAKFVEIYSSA
NLRFVGENNLSANLSFIRENNLSANLSRISTYSKRVCAICRGQPPAKPQICDVGAG
HMLANLEIYSSAKKRPVGEIYSTGTRKRFGEIYSSANLRFVGENSSANLSFIGNNS
SANLSRISTYRKCAVATGGTQPPQSPKRFATPLRLGIMTSTQFPNCLANLRKTYSS
LRLVLPFRDLFVGEKPSLICDVFAGNHDHFAVSSQLSHKFVREIYTKRATETRLG
PVPSSSRNSVAI"
complement(35843..36227)

```

```

/note="contains similarity to non-LTR retroelement reverse
transcriptase
gene_id:K17E7.9"
/codon_start=1
/pseudo
/evidence=not_experimental
join(36811..37599,37734..38069,38621..38818)
/note="gene_id:K17E7.10"
/codon_start=1
/evidence=not_experimental
/product="mitochondrial protein-like; contains similarity
to AAA-type ATPase"
/protein_id="BAB02995.1"
/db_xref="GI:13477082"
/translation="MMLKLIKSMVAITRPIQYLIISYLRIVGPPSLTHNOYTV
IEETSENGRINIVHGATQAYLEDKINLDFEERERDDIYGAKIKMRFFVKNNGNI
PROCELEFDEKEDLDVPSYIPPEVSEKAKIKSKRRLMEHTTSHCCDWETILDH
HSSEFTYMKEDIKRLIDDDIPLISKEDFTKRGKRMRYLLHGLPGAKTSLVA
TAKYINPDYVNIQGVKTDDPTFRLIRVEDSILVIEDIDTSLSEGSVALSQLSL
TPWNSGARVAVITFTNKKERFDQTLCRMEMKTYMGHCCEDEKTLASNYLISHN
DAPRLYPDILRLIDGQAVTPGVVEELMSQDVVALQSLVRYSSSENDHIDDDLP
QIPETRNNSMIDSKPKRSQTIIVEALNVCNTEETGNSNDYSSNKK"
/complement(43355..43976)
/note="gb|AAFI8539.1
gene_id:K17E7.11
similar to unknown protein"
/codon_start=1
/pseudo
/evidence=not_experimental
join(44927..44984,45253..45347,45463..45919,46003..46159,
46227..46504,46590..46873)
/note="gene_id:K17E7.12"
/codon_start=1
/evidence=not_experimental
/product="phyllochelatin synthetase-like protein"
/protein_id="BAB02996.1"
/db_xref="GI:13477083"
/translation="MNTLPSRPSFLILFLCSWNTLNSTEARVADLPYGNITIKDIMS
WTDGQYAVVTITRFOQYRITHEAPWGLQSMWKKKEYTWSWGOATLQDQCSKFGN
IPHCCKTPAIVIDLPGTPYNOQIISNCRGVIISAMADPATAISSFOISGSGSTN
ITVAPRNTILKAPGPGTQGPARKLVKPSRFISADKRKROSILTNINICTYQSFAR
KTPCCVSLSAFNETIVPCPTSCGCGNSOAGTCVDCDPSKIASVAVPAGKNLEP
LLCOTCMCPDIRVHMVHTSYKEKMGCAITPNVNMNYSQANILVQHPNFNLTKL
FSFNKYPLNLYNINDPAMGKIKFYNDPLSQSPGPNVOSSEILFRNDELFEFEKW
APPRRIYNGNDNMVPPDPSYPLNPNASPNATSPVILLITFLSVLLIM"
50023..51531
/note="gb|AAI26945.1
gene_id:K17E7.13
similar to unknown protein"
/codon_start=1
/pseudo
/evidence=not_experimental
join(53764..54696,55011..55204,55298..55562)
/note="emb|CAB71883.1
gene_id:K17E7.14
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB02997.1"
/db_xref="GI:13477084"
/translation="MDSVQNRDIRISLDPVVLVMIISFLSKDNVKTSLSKRMNIC
YEAKNISKESEYVDISVDYKRSRVSFVRLVNWRSVIVISFELGCPGPE
VEIKSLIFRSIROYKYLIVDFSSPFWNTMGLANDPVELPALITYGLTLESLTI
YACMFDPSRFNWGLRTLSIGWFRLEKTESILSKRPLLESISCYDEVKLAQIR
VRELIINDCITPTWYCLINTPRIETFKISGNTVDFKRVNMILKEVLEDFEHEH
DEPRYSKREAGDITSLHNDLSSRTLVCSYLEVISECDNPVMDLMDVQAHLVLE
TLMHPNEFTGMRLLDHCPLNETLFLQCLCKPVPVLOYIDSHTLMENTSENRIIR
TKLIVRGFCNSNNEFYILNLYLPEHGFALEVEYLTPLQLEMPQMAVHGAAMLQ

```

Query Match 6.3%; Score 20; DB 8; Length 81580;  
 Best Local Similarity 100.0%; Pred. No. 7.2;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 176 aaatttgaataatagattgt 195
Db 42912 AAATTTGAAATATGATGCT 42893

RESULT 9
AC105599
LOCUS AC105599
DEFINITION Rattus norvegicus clone CH230-127D2, *** SEQUENCING IN PROGRESS
ACCESSION AC105599
VERSION AC105599.1 GI:18092822
KEYWORDS HTG; HTGS-PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS 1 (bases 1 to 87966)
Mazny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Aisbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dachorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.B., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Hollway,C.,
Hollins,B., Homsl,F., Howard,S., Huber,J., Hulky,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kradovic,J., Kuresh,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Lousaged,L., Losado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., Mcleod,M.P., Meador,M.,
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkwo,S.,
Ogub,M., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N.,
Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,
Stone,H., Tansey,J., Taylor,C., Taylor,T., Tellod,B., Thomas,K.,
Thomas,S., Usmanu,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Walli,R., Wang,S., Ward-Moore,S., Warren,G., Williamson,A., Wleczkyk,R.,
Watlington,S., Williams,G., Williams,A., Wleczkyk,R., Woden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Wenstock,G. and Glibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 87966)
Worley,K.C.
Direct Submission
Submitted (09-JUN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GNHA

```

```

Center clone name: CH230-127D2
----- Summary Statistics
Assembly Program: Phrap; version 0.990329first call to
findphraplist
Consensus quality: 64078 bases at least Q40
Consensus quality: 70281 bases at least Q30
Consensus quality: 75712 bases at least Q20
Estimated insert size: 54379; sum-of-contrigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 0.5x in Q20 bases; sum-of-contrigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank.draft.data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 52 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
3355: contrig of 3355 bp in length
3356
3356: gap of unknown length
3455: gap of unknown length
3456
6405: contrig of 2950 bp in length
6406
6505: gap of unknown length
6506
8439: contrig of 1934 bp in length
8440
8539: gap of unknown length
8540
11712: contrig of 3173 bp in length
11713
11812: gap of unknown length
11813
14902: contrig of 3090 bp in length
14903
15002: gap of unknown length
15003
16825: contrig of 1823 bp in length
16826
16925: gap of unknown length
16926
18687: contrig of 1762 bp in length
18688
18787: gap of unknown length
18788
20481: contrig of 1694 bp in length
20482
20581: gap of unknown length
22234: contrig of 1653 bp in length
22334: gap of unknown length
22335
24360: contrig of 2026 bp in length
24361
24460: gap of unknown length
24461
27456: contrig of 2996 bp in length
27457
27556: gap of unknown length
29429: contrig of 1873 bp in length
29529: gap of unknown length
29530
31544: contrig of 2015 bp in length
31545
31644: gap of unknown length
31645
34443: contrig of 2799 bp in length
34444
34543: gap of unknown length
34544
36482: contrig of 1939 bp in length
36483
36582: gap of unknown length
37738: contrig of 1156 bp in length
37739
37838: gap of unknown length
39324: contrig of 1486 bp in length
39325
39424: gap of unknown length
40701: contrig of 1277 bp in length
40702
40801: gap of unknown length
42655: contrig of 1764 bp in length
42656
42665: gap of unknown length
44320: contrig of 1655 bp in length
44321
44420: gap of unknown length
45953: contrig of 1533 bp in length
45954
46053: gap of unknown length
47550: contrig of 1497 bp in length
47551
47650: gap of unknown length
49656: contrig of 2006 bp in length
49657
49756: gap of unknown length
50789: contrig of 1033 bp in length
50790
50889: gap of unknown length
52357: contrig of 1468 bp in length
52358
52457: gap of unknown length
53677: contrig of 1220 bp in length
53678
53777: gap of unknown length

```

```

* 53778 55154: contrig of 1377 bp in length
* 55155 55254: gap of unknown length
* 55255 56453: contrig of 1199 bp in length
* 56454 56553: gap of unknown length
* 56554 57860: contrig of 1307 bp in length
* 57861 57960: gap of unknown length
* 57961 59018: contrig of 1058 bp in length
* 59019 59118: gap of unknown length
* 59119 60405: contrig of 1287 bp in length
* 60406 60505: gap of unknown length
* 60506 61997: contrig of 1492 bp in length
* 61998 62097: gap of unknown length
* 62098 63125: contrig of 1028 bp in length
* 63126 63325: gap of unknown length
* 63326 64318: contrig of 1093 bp in length
* 64319 64418: gap of unknown length
* 64419 65848: contrig of 1430 bp in length
* 65849 65948: gap of unknown length
* 65949 67048: contrig of 1100 bp in length
* 67049 67148: gap of unknown length
* 67149 68348: contrig of 1200 bp in length
* 68349 68448: gap of unknown length
* 68449 69891: contrig of 1443 bp in length
* 69892 69991: gap of unknown length
* 69992 71020: contrig of 1029 bp in length
* 71021 71120: gap of unknown length
* 71121 72149: contrig of 1029 bp in length
* 72150 72249: gap of unknown length
* 72250 73410: contrig of 1161 bp in length
* 73411 73510: gap of unknown length
* 73511 74552: contrig of 1042 bp in length
* 74552 74652: gap of unknown length
* 74653 75676: contrig of 1024 bp in length
* 75677 75776: gap of unknown length
* 75777 77318: contrig of 1342 bp in length
* 77319 77418: gap of unknown length
* 77419 78720: contrig of 1302 bp in length
* 78721 78820: gap of unknown length
* 78821 79845: contrig of 1025 bp in length
* 79846 79945: gap of unknown length
* 79946 81305: contrig of 1360 bp in length
* 81306 81405: gap of unknown length
* 81406 82699: contrig of 1294 bp in length
* 82700 82799: gap of unknown length
* 82800 84018: contrig of 1219 bp in length
* 84019 84118: gap of unknown length
* 84119 85188: contrig of 1070 bp in length
* 85189 85288: gap of unknown length
* 85289 86830: contrig of 1542 bp in length
* 86831 86930: gap of unknown length
* 86931 87965: contrig of 1036 bp in length.
*
FEATURES
    source          1..87966
                    /organism="Rattus norvegicus"
                    /db_xref="taxon:10116"
                    /clone="CH230-127D2"
BASE COUNT      21580 a 20480 c 19591 g 21063 t 5252 others
ORIGIN
Query Match      6.3%; Score 20; DB 2; Length 87966;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 21 aagaaataactttgagaaa 40
|||||
Db 1669 AAGAAATACTTTTGAGAAA 1688
|||||
RESULT 10
AC027033/C 98017 bp DNA linear PLN 11-OCT-2000
LOCUS AC027033
DEFINITION Arabidopsis thaliana chromosome 1 BAC F21N10 genomic sequence,

```

complete sequence.  
 AC027033 GI:10092252  
 VERSION AC027033.3  
 KEYWORDS HTG.  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 98017)  
 Lin, X., Kaul, S., Town, C.D., Benito, M.-I., Creasy, T.H., Haas, B.J., Barnstead, M.E., Bowman, C.L., White, O., Niernan, W.C., and Fraser, C.M. Arabidopsis thaliana chromosome 1 BAC F21N10 genomic sequence  
 Unpublished  
 2 (bases 1 to 98017)  
 Lin, X. and Kaul, S.  
 Direct Submission  
 Submitted (26-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org  
 3 (bases 1 to 98017)  
 Town, C.D. and Kaul, S.  
 Direct Submission  
 Submitted (12-SEP-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org  
 On Sep 12, 2000 this sequence version replaced gi:7709919.  
 Address all correspondence to: atetigr.org

COMMENT  
 BAC clone F21N10 is from Arabidopsis thaliana chromosome 1  
 The orientation of the sequence is from SP6 to T7 end of the BAC clone.  
 Genes were identified by a combination of several methods: Gene prediction programs including Genscan+ (Chris Burge, <http://ccr-081.mt.edu/GENSCAN.html>), GeneMarkHM (Mark Borodovsky, <http://genemark.biology.gatech.edu/GeneMark/>), Glimmer (a variant of Glimmer, see Mihaela Perlea, <http://www.tigr.org/softlab/glimmer.htm>), and GeneSplicer (Mihaela Perlea and Steven Salzberg, contact [mpertea@tigr.org](mailto:mpertea@tigr.org)), searches of the complete sequence against a peptide database and the plant EST database at TIGR (<http://www.tigr.org/cdb/cgi.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).  
 Location/Qualifiers  
 1..98017  
 /organism="Arabidopsis thaliana"  
 /cultivar="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="1"  
 /clone="F21N10"  
 1184..5698  
 /gene="F21N10.12"  
 /note="pseudogene, polyprotein"  
 /pseudo  
 1368..5694  
 /rpt\_family="ATCOP1A16|ATCOP1A16I Internal region of ATCOP1A16 copia-like LTR-retrotransposon."  
 1397..3853  
 /rpt\_family="ATCOP1A11|ATCOP1A11 Internal region of ATCOP1A1 LTR-retrotransposon."  
 1407..1704  
 /rpt\_family="RE1|RE1 Arabidopsis thaliana retroelement - a consensus."  
 3037..3529  
 /rpt\_family="ATCOP1A11|ATCOP1A11 Internal region of

repeat\_region ATCOP1A11 LTR-retrotransposon."  
 3037..3795  
 /rpt\_family="ATCOP1A15I|ATCOP1A15I Internal region of ATCOP1A15 copia-like LTR-retrotransposon."  
 3124..3348  
 /rpt\_family="ATCOP1A17I|ATCOP1A17I Internal region of ATCOP1A17 copia-like LTR-retrotransposon."  
 3407..3585  
 /rpt\_family="ATCOP1A24I|ATCOP1A24I Internal region of ATCOP1A24 copia-like LTR-retrotransposon."  
 3628..3797  
 /rpt\_family="ATCOP1A11I|ATCOP1A11I Internal region of ATCOP1A11 LTR-retrotransposon."  
 4256..4543  
 /rpt\_family="ATCOP1A15I|ATCOP1A15I Internal region of ATCOP1A15 copia-like LTR-retrotransposon."  
 4437..4629  
 /rpt\_family="ATCOP1A11I|ATCOP1A11I Internal region of ATCOP1A1 LTR-retrotransposon."  
 4662..4969  
 /rpt\_family="ATCOP1A11I|ATCOP1A11 Internal region of ATCOP1A1 LTR-retrotransposon."  
 7315..7414  
 /rpt\_family="ARNOLD3|ARNOLD3 Autonomous DNA-transposon ARNOLD3."  
 7315..7416  
 /rpt\_family="ARNOLD4|ARNOLD4 Autonomous DNA-transposon ARNOLD4."  
 7328..7416  
 /rpt\_family="ARNOLD2|ARNOLD2 Autonomous DNA-transposon ARNOLD2."  
 7328..7416  
 /rpt\_family="ARNOLD1|ARNOLD1 Autonomous DNA-transposon ARNOLD1."  
 8053..10391  
 /rpt\_family="ARNOLD3|ARNOLD3 Autonomous DNA-transposon ARNOLD3."  
 8053..10391  
 /rpt\_family="ARNOLD4|ARNOLD4 Autonomous DNA-transposon ARNOLD4."  
 8071..10391  
 /rpt\_family="ARNOLD1|ARNOLD1 Autonomous DNA-transposon ARNOLD1."  
 8785..8913  
 /rpt\_family="ARNOLD4|ARNOLD4 Autonomous DNA-transposon ARNOLD4."  
 8785..8913  
 /rpt\_family="ARNOLD3|ARNOLD3 Autonomous DNA-transposon ARNOLD3."  
 8803..8913  
 /rpt\_family="ARNOLD1|ARNOLD1 Autonomous DNA-transposon ARNOLD1."  
 8803..8913  
 /rpt\_family="ARNOLD2|ARNOLD2 Autonomous DNA-transposon ARNOLD2."  
 8830..8913  
 /rpt\_family="ARNOLDY2|ARNOLDY2 Non-autonomous DNA transposon ARNOLDY2 - a consensus sequence."  
 9506..9634  
 /rpt\_family="ARNOLD3|ARNOLD3 Autonomous DNA-transposon ARNOLD3."  
 9506..9634  
 /rpt\_family="ARNOLD4|ARNOLD4 Autonomous DNA-transposon ARNOLD4."  
 9525..9634  
 /rpt\_family="ARNOLD1|ARNOLD1 Autonomous DNA-transposon ARNOLD1."  
 9525..9634  
 /rpt\_family="ARNOLD2|ARNOLD2 Autonomous DNA-transposon ARNOLD2."  
 10984..11112  
 /rpt\_family="ARNOLD3|ARNOLD3 Autonomous DNA-transposon ARNOLD3."

```

repeat_region      10984..11112
                    /rpt_family="ARNOLD4|ARNOLD4 Autonomously DNA-transposon
ARNOLD4."
repeat_region      11001..11112
                    /rpt_family="ARNOLD1|ARNOLD1 Autonomously DNA-transposon
ARNOLD1."
repeat_region      11001..11112
                    /rpt_family="ARNOLD2|ARNOLD2 Autonomously DNA-transposon
ARNOLD2."
repeat_region      complement(11870..12010)
                    /rpt_family="ATREP5|ATREP5 An AT-rich repetitive sequence
                    a consensus."
repeat_region      12462..12590
                    /rpt_family="ARNOLD4|ARNOLD4 Autonomously DNA-transposon
ARNOLD4."
repeat_region      12462..12590
                    /rpt_family="ARNOLD3|ARNOLD3 Autonomously DNA-transposon
ARNOLD3."
repeat_region      12479..12590
                    /rpt_family="ARNOLD1|ARNOLD1 Autonomously DNA-transposon
ARNOLD1."
repeat_region      12479..12590
                    /rpt_family="ARNOLD2|ARNOLD2 Autonomously DNA-transposon
ARNOLD2."
repeat_region      13255..13383
                    /rpt_family="ARNOLD4|ARNOLD4 Autonomously DNA-transposon
ARNOLD4."
repeat_region      13255..13383
                    /rpt_family="ARNOLD3|ARNOLD3 Autonomously DNA-transposon
ARNOLD3."
repeat_region      13273..13383
                    /rpt_family="ARNOLD1|ARNOLD1 Autonomously DNA-transposon
ARNOLD1."
repeat_region      13273..13383
                    /rpt_family="ARNOLD2|ARNOLD2 Autonomously DNA-transposon
ARNOLD2."
repeat_region      13975..14103
                    /rpt_family="ARNOLD3|ARNOLD3 Autonomously DNA-transposon
ARNOLD3."
repeat_region      13975..14103
                    /rpt_family="ARNOLD4|ARNOLD4 Autonomously DNA-transposon
ARNOLD4."
repeat_region      13975..14103
                    /rpt_family="ARNOLD2|ARNOLD2 Autonomously DNA-transposon
ARNOLD2."
repeat_region      13993..14103
                    /rpt_family="ARNOLD2|ARNOLD2 Autonomously DNA-transposon
ARNOLD2."
repeat_region      13993..14103
                    /rpt_family="ARNOLD1|ARNOLD1 Autonomously DNA-transposon
ARNOLD1."
repeat_region      15628..15837
                    /rpt_family="centromeric repeat T5H22 31150 to 31597"
repeat_region      16107..16191
                    /rpt_family="centromeric repeat T5H22 31150 to 31597"

Query Match      6.3%; Score 20; DB 8; Length 98017;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 aaatttgaataatgattgt 195
Db 29803 AAATTTGAATAATGATTGT 29784

```

```

REFERENCE
AUTHORS
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 109981)
Muzny,D.M., Adams,C., Adio-Oduola,B., All-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbarella,J.,
Benton,J., Bimage,K., Blankenburg,K., Bonini,D., Bouck,J.,
Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N., Buhay,C.,
Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douhaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C.,
Hollins,B., Homsli,F., Howard,S., Huber,J., Huliy,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korval,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Maxhiney,E., McLeod,M.P., Meador,M.,
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokweto,S.,
Ogulu,M., Okunolu,G., Oregunye,N., Oviado,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
Sisson,I., Sodergren,E., Sonalite,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Verd,V., Villalona,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,A., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wlecyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 109981)
Worley,K.C.
Direct Submission
Submitted (28-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15799472.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: GGVW
Center clone name: CH230-91G16
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 67454 bases at least Q40
Consensus quality: 77134 bases at least Q30
Consensus quality: 85466 bases at least Q20
Estimated insert size: 72753; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 0.7x in Q20 bases; sum-of-contigs estimation

-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 59 contigs. The true order of the pieces

```

\* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence.  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1  
 2162: contig of 2162 bp in length  
 2163 2262: gap of unknown length  
 2263 4905: contig of 2643 bp in length  
 4906 5005: gap of unknown length  
 5006 7945: contig of 2940 bp in length  
 7946 8046: gap of unknown length  
 8046 9668: contig of 1623 bp in length  
 9669 9768: gap of unknown length  
 9769 12307: contig of 2539 bp in length  
 12308 12407: gap of unknown length  
 12408 14880: contig of 2473 bp in length  
 14881 14980: gap of unknown length  
 14981 18246: contig of 3286 bp in length  
 18247 18346: gap of unknown length  
 18347 20802: contig of 2456 bp in length  
 20803 20902: gap of unknown length  
 20903 23429: contig of 2527 bp in length  
 23430 23529: gap of unknown length  
 23530 26244: contig of 2715 bp in length  
 26245 26344: gap of unknown length  
 26345 28233: contig of 1889 bp in length  
 28234 28333: gap of unknown length  
 28334 30557: contig of 2224 bp in length  
 30558 30657: gap of unknown length  
 30658 32242: contig of 1585 bp in length  
 32243 34175: contig of 1833 bp in length  
 34176 34275: gap of unknown length  
 34276 37250 37250: contig of 2974 bp in length  
 37250 39006: contig of 1657 bp in length  
 39007 39106: gap of unknown length  
 39107 40848: contig of 1742 bp in length  
 40849 40948: gap of unknown length  
 40949 42684: contig of 1736 bp in length  
 42685 42784: gap of unknown length  
 42785 45207: contig of 2423 bp in length  
 45208 45307: gap of unknown length  
 45308 46998: contig of 1691 bp in length  
 46999 47099: gap of unknown length  
 47099 48373: contig of 1275 bp in length  
 48374 48473: gap of unknown length  
 48474 49758: contig of 1285 bp in length  
 49759 49859: gap of unknown length  
 49859 51552: contig of 1694 bp in length  
 51553 51652: gap of unknown length  
 51653 53239: contig of 1587 bp in length  
 53240 53339: gap of unknown length  
 53340 54649: contig of 1310 bp in length  
 54650 54749: gap of unknown length  
 54750 56982: contig of 2233 bp in length  
 56983 57082: gap of unknown length  
 57083 58246: contig of 1164 bp in length  
 58247 58346: gap of unknown length  
 58347 60579: contig of 2133 bp in length  
 60580 60779: gap of unknown length  
 60780 62706: contig of 2127 bp in length  
 62707 62806: gap of unknown length  
 62807 64618: contig of 1812 bp in length  
 64619 64718: gap of unknown length  
 64719 66415: contig of 1697 bp in length  
 66416 66515: gap of unknown length  
 66516 67617: contig of 1102 bp in length  
 67618 69472: contig of 1755 bp in length  
 69473 69572: gap of unknown length  
 69573 70847: contig of 1275 bp in length

\* 70848 70947: gap of unknown length  
 \* 70948 72752: contig of 1805 bp in length  
 \* 72753 72852: gap of unknown length  
 \* 72853 74624: contig of 1772 bp in length  
 \* 74625 74724: gap of unknown length  
 \* 74725 76802: contig of 2078 bp in length  
 \* 76803 76902: gap of unknown length  
 \* 76903 78528: contig of 1626 bp in length  
 \* 78529 78628: gap of unknown length  
 \* 78629 80110: contig of 1482 bp in length  
 \* 80111 80210: gap of unknown length  
 \* 80211 82357: contig of 2147 bp in length  
 \* 82358 82457: gap of unknown length  
 \* 82458 83925: contig of 1468 bp in length  
 \* 83926 84025: gap of unknown length  
 \* 84026 85056: contig of 1031 bp in length  
 \* 85057 85156: gap of unknown length  
 \* 85157 86839: contig of 1683 bp in length  
 \* 86840 86939: gap of unknown length  
 \* 86940 88378: contig of 1439 bp in length  
 \* 88379 88478: gap of unknown length  
 \* 88479 89569: contig of 1091 bp in length  
 \* 89570 89669: gap of unknown length  
 \* 89670 91105: contig of 1436 bp in length  
 \* 91106 91205: gap of unknown length  
 \* 91206 92755: contig of 1550 bp in length  
 \* 92756 92856: gap of unknown length  
 \* 92857 94275: contig of 1420 bp in length  
 \* 94276 94375: gap of unknown length  
 \* 94376 96089: contig of 1714 bp in length  
 \* 96090 96189: gap of unknown length  
 \* 96190 97291: contig of 1102 bp in length  
 \* 97292 97391: gap of unknown length  
 \* 97392 98457: contig of 1066 bp in length  
 \* 98458 98557: gap of unknown length  
 \* 98559 99593: contig of 1036 bp in length  
 \* 99594 99693: gap of unknown length  
 \* 99694 100830: contig of 1137 bp in length  
 \* 100831 100930: gap of unknown length  
 \* 100931 102463: contig of 1533 bp in length  
 \* 102464 102563: gap of unknown length  
 \* 102564 104121: contig of 1558 bp in length  
 \* 104122 104221: gap of unknown length  
 \* 104222 105573: contig of 1352 bp in length  
 \* 105574 105673: gap of unknown length

Query Match 5.3% Score 20: DB 2: Length 109981;  
 Best Local Similarity 100.0% Pred. No. 7;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 44 aacaaatgaatcagttct 63  
 DB 93606 AACAAATGAATCAGTTCT 93587

RESULT 12  
 LOCUS AC008271 111159 bp DNA linear PRI 09-MAY-2001  
 DEFINITION Homo sapiens BAC clone RP11-123E16 from 2, complete sequence.  
 AC008271  
 VERSION AC008271.6 GI:11120938  
 KEYWORDS HTG.

SOURCE  
 ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS Sulston, D.E. and Waterston, R.

Toward a complete human genome sequence  
 Genome Res. 8 (11), 1097-1108 (1998)

MEDLINE

REFERENCE 2 (bases 1 to 111159)  
 Haakenson, B., Stoneking, T., Elliott, G. and Phillips, A.

TITLE The sequence of Homo sapiens BAC clone RP11-123E16  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 111159)  
AUTHORS Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (31-JUL-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
REFERENCE 4 (bases 1 to 111159)  
AUTHORS Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (08-NOV-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
REFERENCE 5 (bases 1 to 111159)  
AUTHORS Waterston, R.  
TITLE Direct Submission  
JOURNAL Submitted (09-MAY-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
COMMENT On Nov 8, 2000 this sequence version replaced gi:9838078.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc  
Contact: saplens@watson.wustl.edu  
----- Summary Statistics  
Center project name: H\_NH0123E16  
-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:  
The RPCT-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, E., Tateo, M., Catanese, J.J., and de Jong, P.U. (1996) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)  
VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the left is RP11-422A6, 200 bp overlap; the clone sequenced to the right is RP11-120J4. Actual start of this clone is at base position 84248 of RP11-422A6; actual end is at base position 111159 of RP11-123E16.

The sequence RP11-123E16 contains a dinucleotide (GA) repeat from base position 3114 to 3320. There is single subclone coverage and the fidelity of sequence cannot be guaranteed. The region corresponds to restriction digest information: BamHI band size 6715 real and 6686 in silico; and HindIII band size 3806 real and 3801 in silico.

FEATURES  
source  
1. 111159  
Location/Qualifiers  
/organism="Homo sapiens"

/db\_xref="taxon:9606"  
/chromosome="2"  
/map="2"  
/clone="RP11-123E16"  
/clone\_11b="RPCT-11"  
158. .401  
/rpt\_family="MIR"  
648. .950  
repeat\_region  
/rpt\_family="L1"  
1095. .1220  
repeat\_region  
/rpt\_family="MALR"  
1272. .1424  
repeat\_region  
/rpt\_family="MERL\_type"  
1437. .1558  
repeat\_region  
/rpt\_family="MERL\_type"  
1570. .1732  
repeat\_region  
/rpt\_family="MALR"  
1758. .1920  
repeat\_region  
/rpt\_family="L2"  
1993. .3207  
repeat\_region  
/rpt\_family="SVA"  
2228. .3046  
misc\_feature  
/note="CpG\_Island (%GC=68.4, o/e=0.70, #CPGs=59)"  
4155. .4564  
repeat\_region  
/rpt\_family="MALR"  
6437. .6553  
repeat\_region  
/rpt\_family="MIR"  
6581. .6670  
repeat\_region  
/rpt\_family="L2"  
7520. .7664  
repeat\_region  
/rpt\_family="MALR"  
7702. .7967  
repeat\_region  
/rpt\_family="Alu"  
8102. .8321  
repeat\_region  
/rpt\_family="Alu"  
8337. .8569  
repeat\_region  
/rpt\_family="MALR"  
8725. .9020  
repeat\_region  
/rpt\_family="MALR"  
9115. .9184  
repeat\_region  
/rpt\_family="MALR"  
9231. .9282  
repeat\_region  
/rpt\_family="Alu"  
9244. .9395  
repeat\_region  
/rpt\_family="MALR"  
9397. .9773  
repeat\_region  
/rpt\_family="MALR"  
10212. .10420  
repeat\_region  
/rpt\_family="MIR"  
11260. .11350  
repeat\_region  
/rpt\_family="L2"  
11305. .11587  
repeat\_region  
/rpt\_family="MIR"  
12341. .12645  
repeat\_region  
/rpt\_family="Alu"  
13003. .13424  
repeat\_region  
/rpt\_family="ACHOB0"  
13645. .13700  
repeat\_region  
/rpt\_family="MIR"  
14661. .14743  
repeat\_region  
/rpt\_family="MIR"  
15183. .15284  
repeat\_region  
/rpt\_family="Alu"  
15589. .15681  
repeat\_region  
/rpt\_family="Alu"  
15702. .16189  
repeat\_region  
/rpt\_family="L1"  
16755. .16896  
repeat\_region  
/rpt\_family="L2"  
17055. .17236  
repeat\_region  
/rpt\_family="L2"  
18001. .18245  
repeat\_region  
/rpt\_family="ERV1"

```
repeat_region      18249. 18319  
/rpt_family="ERV1"  
repeat_region      18324. 19402  
/rpt_family="ERV1"  
repeat_region      19507. 19757  
/rpt_family="L1"  
repeat_region      20001. 20324  
/rpt_family="Alu"  
repeat_region      21240. 21343  
/rpt_family="MIR"  
repeat_region      21356. 21529  
/rpt_family="MALR"  
repeat_region      21530. 22306  
/rpt_family="L1"  
repeat_region      22326. 22660  
/rpt_family="MALR"  
repeat_region      22783. 23121  
/rpt_family="L2"  
repeat_region      23450. 23536  
/rpt_family="MIR"  
repeat_region      23590. 24280  
/rpt_family="ERV1"  
repeat_region      24424. 24593  
/rpt_family="Alu"  
repeat_region      24648. 25136  
/rpt_family="ERV1"  
repeat_region      25137. 25493  
/rpt_family="MALR"  
repeat_region      25494. 25534  
/rpt_family="ERV1"  
repeat_region      27117. 27389  
/rpt_family="L2"  
repeat_region      27552. 28004  
/rpt_family="MALR"  
repeat_region      28142. 28278  
/rpt_family="L2"  
repeat_region      28349. 28505  
/rpt_family="L2"
```

```
Query Match      6.38; Score 20; DB 9; Length 111159;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY      74 agtatactcttagatct 93
      |||
Db 73852 AGTATATCTCTTAGAGATCT 73833

```

Db 73852 AGTATATCTCTTAGAGATCT 73833

RESULT	13		
AC105629			
LOCUS			
DEFINITION	AC105629	114927 bp	DNA
	<i>Rattus norvegicus</i> clone CH230-13J4, ***		linear
	***, 60 unordered pieces.		SEQUENCING IN PROGRESS
ACCESSION	AC105629		
VERSION	AC105629.1	GI:18092852	
KEYWORDS	HTG; HTGS_PHASE1.		
SOURCE	Norway rat.		
ORGANISM	<i>Rattus norvegicus</i>		

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

Elhaj, C., Ascotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodson, A., Hogues, M., Holloway, C., Hollins, B., Homsi, F., Howard, S., Huber, J., Huliy, S., Hume, J., Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Krelli, S., Rhan, U., King, L., Korvah, J., Lewis, L., Liu, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Louised, H., Lozado, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, Y., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhinney, E., Meleod, M. P., Meador, M., Mel, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabhat, K., Morgan, A., Morris, S., Moser, M., Neal, D., Newton, J., Nektson, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nkwenkwo, S., Ogburn, M., Okunolu, G., Orgunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L., Qulies, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I., Sodergren, E., Sonaile, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svaitek, A., Tabori, P., Tamerisa, A., Tang, H., Tansey, A., Taylor, C., Taylor, T., Telford, B., Thomas, S., Thomas, S., Usmani, K., Vasquez, L., Vora, V., Villalton, D., Vinson, R., Wall, R., Wang, S., Ward, Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wood, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 114927)  
Morley, K. C.

Direct Submission  
Submitted (09-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

----- Genome Center -----  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
Project Information  
Center project name: GKNB  
Center clone name: CH230-13124  
----- Summary Statistics -----  
Assembly program: Phrap; version 0.990329first call to findPiraplist  
Consensus quality: 92543 bases at least Q40  
Consensus quality: 99732 bases at least Q40  
Consensus quality: 104520 bases at least Q20  
Estimated insert size: 74569; sum-of-coverage estimation  
Quality coverage: 0x in Q20 bases; agarose-ip estimation  
Quality coverage: 1x in Q20 bases; sum-of-coverage estimation

\* NOTE: Estimated insert size may differ from sequence length  
(see <http://www.hgsc.bcm.tmc.edu/docs/genbank.draft.data.html>)  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 60 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1	4013: contig of 4013 bp in length
4014	4113: gap of unknown length
4114	8452: contig of 4339 bp in length
8453	8453: gap of unknown length
8553	11386: contig of 2834 bp in length
11387	11486: gap of unknown length
11487	14815: contig of 3329 bp in length
14816	14915: gap of unknown length
14916	18385: contig of 3670 bp in length



```

* 18586 18685: gap of unknown length
* 18686 22719: contig of 4034 bp in length
* 22720 22819: gap of unknown length
* 22820 25644: contig of 2825 bp in length
* 25645 25744: gap of unknown length
* 25745 27426: contig of 1682 bp in length
* 27427 27526: gap of unknown length
* 27527 28729: contig of 1203 bp in length
* 28730 28829: gap of unknown length
* 28830 30452: contig of 1623 bp in length
* 30453 30552: gap of unknown length
* 30553 33530: contig of 2978 bp in length
* 33531 33630: gap of unknown length
* 33631 36409: contig of 2779 bp in length
* 36410 36509: gap of unknown length
* 36510 37569: contig of 1060 bp in length
* 37570 37669: gap of unknown length
* 37670 38852: contig of 1183 bp in length
* 38853 38952: gap of unknown length
* 38953 40293: contig of 1341 bp in length
* 40294 40393: gap of unknown length
* 40394 42174: contig of 1781 bp in length
* 42175 42274: gap of unknown length
* 42275 44226: contig of 1952 bp in length
* 44227 44326: gap of unknown length
* 44327 45602: contig of 1276 bp in length
* 45603 45702: gap of unknown length
* 45703 46336: contig of 1234 bp in length
* 46337 47036: gap of unknown length
* 47037 49538: contig of 2502 bp in length
* 49539 51322: contig of 1684 bp in length
* 51323 51422: gap of unknown length
* 51423 54203: contig of 2781 bp in length
* 54204 54503: gap of unknown length
* 54504 56244: contig of 1941 bp in length
* 56245 56344: gap of unknown length
* 56345 58676: contig of 2332 bp in length
* 58677 58776: gap of unknown length
* 58777 60187: contig of 1411 bp in length
* 60188 60287: gap of unknown length
* 60288 62280: contig of 1993 bp in length
* 62281 62380: gap of unknown length
* 62381 64324: contig of 1944 bp in length
* 64325 64424: gap of unknown length
* 64425 66005: contig of 1581 bp in length
* 66006 66105: gap of unknown length
* 66106 68093: contig of 1988 bp in length
* 68094 68193: gap of unknown length
* 68194 69970: contig of 1777 bp in length
* 69971 70070: gap of unknown length
* 70071 71599: contig of 1529 bp in length
* 71600 71699: gap of unknown length
* 71700 72859: contig of 1160 bp in length
* 72860 72959: gap of unknown length
* 72960 74333: contig of 1374 bp in length
* 74334 74433: gap of unknown length
* 74434 76033: contig of 1600 bp in length
* 76034 76133: gap of unknown length
* 76134 78122: contig of 1989 bp in length
* 78123 78222: gap of unknown length
* 78223 80015: contig of 1793 bp in length
* 80016 80115: gap of unknown length
* 80116 81418: contig of 1303 bp in length
* 81419 81518: gap of unknown length
* 81519 83009: contig of 1491 bp in length
* 83010 83109: gap of unknown length
* 83110 84715: contig of 1606 bp in length
* 84716 84815: gap of unknown length
* 84816 86128: contig of 1313 bp in length
* 86129 86228: gap of unknown length
* 86229 88570: contig of 2342 bp in length
* 88571 88670: gap of unknown length

```

```

* 88671 89768: contig of 1098 bp in length
* 89769 89868: gap of unknown length
* 89869 90939: contig of 1071 bp in length
* 90940 91039: gap of unknown length
* 91040 92547: contig of 1508 bp in length
* 92548 92647: gap of unknown length
* 92648 94041: contig of 1394 bp in length
* 94042 94141: gap of unknown length
* 94142 95517: contig of 1376 bp in length
* 95518 95617: gap of unknown length
* 95618 96690: contig of 1073 bp in length
* 96691 96790: gap of unknown length
* 96791 97926: contig of 1136 bp in length
* 97927 98026: gap of unknown length
* 98027 99504: contig of 1478 bp in length
* 99505 99604: gap of unknown length
* 99605 101254: contig of 1650 bp in length
* 101255 101354: gap of unknown length
* 101355 103141: contig of 1787 bp in length
* 103142 103241: gap of unknown length
* 103242 104438: contig of 1197 bp in length
* 104439 104538: gap of unknown length
* 104539 105718: contig of 1180 bp in length
* 105719 105818: gap of unknown length
* 105819 106835: contig of 1017 bp in length
* 106836 106935: gap of unknown length
* 106936 108585: contig of 1650 bp in length
* 108586 108685: gap of unknown length
* 108686 109842: contig of 1157 bp in length
* 109843 109942: gap of unknown length
* 109943 111491: contig of 1549 bp in length

Query Match      6.3%; Score 20; DB 2; Length 114927;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      21 aagaataacttttgagaa 40
Db 96146 AAGAAATACCTTTTGAGAA 96165
|||||
|||||

RESULT 14
AC094962
LOCUS
DEFINITION Rattus norvegicus clone CH230-6L19, *** SEQUENCING IN PROGRESS ***,
AC094962
ACCESSION AC094962
VERSION AC094962.2 GI:17941763
KEYWORDS HTG; HTGS; PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 115948)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Alf-oman,F.R., Allen,C.,
Albrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimarge,K., Blankenburg,K., Bonin,D., Bouck,J.,
Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C.,
Butcher,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Doultwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hughes,M., Holloway,C.,
Hollins,B., Homs,J.F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,

```

Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovari, C., Kralovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L., Li, J., Li, Z., Lichtharge, O., Lieu, C., Liu, J., Liu, W., Loulseg, H., Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M. P., Meador, M., Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabati, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nockenkwo, S., Oguh, M., Okunolu, G., Orangun, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojudo, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoshitari, N., Stone, H., Sutton, A., Swatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, P., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wallington, S., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wellington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 115948)  
Worley, K.C.

Submitted (15-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Dec 20, 2001 this sequence version replaced gi:15624799.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information  
Center project name: GRUX  
Center clone name: CH230-6L19  
----- Summary Statistics  
findPhrapList  
Assembly program: Phrap; version 0.990329first call to  
Consensus quality: 81103 bases at least Q40  
Consensus quality: 94888 bases at least Q30  
Consensus quality: 104411 bases at least Q20  
Estimated insert size: 51502; sum-of-coverage estimation  
Quality coverage: 0x in Q20 bases; agarose-IP estimation  
Quality coverage: 0.5x in Q20 bases; sum-of-coverage estimation

----- NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
NOTE: This is a 'working draft' sequence. It currently  
\* consists of 71 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1	1010:	contig of 1010 bp in length
1011	1110:	gap of unknown length
1111	4665:	contig of 3555 bp in length
4666	4765:	gap of unknown length
4766	8651:	contig of 3886 bp in length
8652	8751:	gap of unknown length
8752	11403:	contig of 2652 bp in length
11404	11503:	gap of unknown length
11504	13360:	contig of 1857 bp in length
13361	13460:	gap of unknown length
13461	15610:	contig of 2150 bp in length
15611	15710:	gap of unknown length
15711	18255:	contig of 2345 bp in length
18256	18355:	gap of unknown length
18356	19947:	contig of 1592 bp in length

19948	20047:	gap of unknown length
20048	21200:	contig of 1153 bp in length
21201	21300:	gap of unknown length
21301	23457:	contig of 2157 bp in length
23458	23557:	gap of unknown length
23558	26085:	contig of 2528 bp in length
26086	26185:	gap of unknown length
26186	27472:	contig of 1287 bp in length
27473	27572:	gap of unknown length
27573	29081:	contig of 1509 bp in length
29082	29181:	gap of unknown length
29182	30555:	contig of 1374 bp in length
30556	30655:	gap of unknown length
30656	32892:	contig of 2237 bp in length
32893	32992:	gap of unknown length
32993	34484:	contig of 1492 bp in length
34485	34584:	gap of unknown length
34585	36413:	contig of 1829 bp in length
36414	36513:	gap of unknown length
36514	38275:	contig of 1762 bp in length
38276	38375:	gap of unknown length
38376	39975:	contig of 1600 bp in length
39976	40076:	gap of unknown length
40076	41487:	contig of 1412 bp in length
41488	41587:	gap of unknown length
41588	42910:	contig of 1323 bp in length
42911	43010:	gap of unknown length
43011	44419:	contig of 1409 bp in length
44420	44519:	gap of unknown length
44520	45673:	contig of 1154 bp in length
45674	45773:	gap of unknown length
45774	47130:	contig of 1357 bp in length
47131	47230:	gap of unknown length
47231	48323:	contig of 1633 bp in length
48324	49023:	gap of unknown length
49024	50794:	contig of 1771 bp in length
50795	50894:	gap of unknown length
50895	52255:	contig of 1361 bp in length
52256	52355:	gap of unknown length
52356	53507:	contig of 1152 bp in length
53508	53607:	gap of unknown length
53608	54936:	contig of 1229 bp in length
54937	54936:	gap of unknown length
54938	56428:	contig of 1492 bp in length
56429	56528:	gap of unknown length
56529	57735:	contig of 1207 bp in length
57736	57835:	gap of unknown length
57836	58872:	contig of 1037 bp in length
58873	58972:	gap of unknown length
58973	60518:	contig of 1546 bp in length
60519	60618:	gap of unknown length
60619	61655:	contig of 1037 bp in length
61656	61755:	gap of unknown length
61756	63254:	contig of 1499 bp in length
63255	63354:	gap of unknown length
63355	64932:	contig of 1578 bp in length
64933	65032:	gap of unknown length
65033	65962:	contig of 1930 bp in length
65963	67062:	gap of unknown length
67063	68267:	contig of 1205 bp in length
68268	68367:	gap of unknown length
68368	69887:	contig of 1520 bp in length
69888	71612:	gap of unknown length
69889	71712:	contig of 1625 bp in length
71713	73377:	gap of unknown length
73378	73477:	contig of 1665 bp in length
73478	75473:	gap of unknown length
75474	75573:	contig of 1996 bp in length
75574	76953:	gap of unknown length
76954	77053:	contig of 1380 bp in length
77054	78130:	gap of unknown length
78131	78230:	contig of 1077 bp in length

COMMENT

\* 78231 79584: contig of 1354 bp in length  
\* 79685 79684: gap of unknown length  
\* 79685 80811: contig of 1127 bp in length  
\* 80812 80911: gap of unknown length  
\* 80812 82052: contig of 1141 bp in length  
\* 82052 82152: gap of unknown length  
\* 82153 83643: contig of 1491 bp in length  
\* 83644 83743: gap of unknown length  
\* 83744 85237: contig of 1484 bp in length  
\* 85238 85328: gap of unknown length  
\* 85328 86706: contig of 1379 bp in length  
\* 86707 86806: gap of unknown length  
\* 86807 87892: contig of 1086 bp in length  
\* 87893 87992: gap of unknown length  
\* 87993 89432: contig of 1440 bp in length  
\* 89433 89532: gap of unknown length  
\* 89533 90625: contig of 1093 bp in length  
\* 90626 90725: gap of unknown length  
\* 90726 91896: contig of 1171 bp in length  
\* 91897 91996: gap of unknown length  
\* 91997 93048: contig of 1052 bp in length  
\* 93049 93148: gap of unknown length  
\* 93149 94569: contig of 1421 bp in length  
\* 94570 94669: gap of unknown length

Query Match 6.3%; Score 20; DB 2; Length 115948;  
Best Local Similarity 100.0%; Pred. No. 7;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 258 acagagatcattacttgcga 277  
|||||  
Db 70157 ACAGAGATCATTTACTTGCA 70176

RESULT 15  
HSJ100419/c 117757 bp DNA linear PRI 15-MAR-2001  
LOCUS  
DEFINITION  
Human DNA sequence from clone RP5-100419 on chromosome 20 Contains  
ESTs, STSS and GSSs, complete sequence.  
ACCESSION  
AL121912  
VERSION  
AL121912.18 GI:7018383  
KEYWORDS  
HTG.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
1 (bases 1 to 117757)  
Direct Submission  
Submitted (09-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
COMMENT  
On Feb 21, 2000 this sequence version replaced gi:6983194.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information  
on the WORMPEP database can be found at  
http://www.sanger.ac.uk/Projects/Celegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 20, constructed by the Sanger Centre Chromosome 20  
Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr20  
This sequence is the entire insert of clone RP5-100419 The true  
left end of clone RP11-500011 is at 24031 in this sequence. This  
sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=

FEATURES  
source

30): an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. RP5-100419 is from  
the library RPc1-5 constructed by the group of Pieter de Jong. For  
further details see  
http://www.choil.org/bacpac/home.htm  
VECTOR: pcypac2.  
Location/Qualifiers  
1..117757  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="20"  
/clone\_1lb="RPc1-5"  
265..312  
/note="24 copies 2 mer aa 728 conserved"  
562..692  
/note="MIR repeat: matches 51..195 of consensus"  
826..913  
/note="L1R16C repeat: matches 272..354 of consensus"  
914..1280  
/note="THE1B repeat: matches 1..364 of consensus"  
1291..1461  
/note="L1R16C repeat: matches 81..272 of consensus"  
2025..2173  
/note="L12 repeat: matches 2344..2492 of consensus"  
2213..2294  
/note="MIR repeat: matches 85..169 of consensus"  
2295..2422  
/note="MER3 repeat: matches 1..128 of consensus"  
2439..2808  
/note="THE1B repeat: matches 1..364 of consensus"  
2913..2975  
/note="MER3 repeat: matches 145..207 of consensus"  
4751..5056  
/note="AluY repeat: matches 2..296 of consensus"  
5841..6068  
/note="MER46A repeat: matches 1..236 of consensus"  
6078..6129  
/note="26 copies 2 mer tt 848 conserved"  
6133..6977  
/note="L1PA3 repeat: matches 5286..6146 of consensus"  
6978..7495  
/note="L1PA3 repeat: matches 4747..5264 of consensus"  
7506..8207  
/note="L1MB3 repeat: matches 5476..6180 of consensus"  
complement(8496..8966)  
/note="match: GSS: Em:AQ487745"  
8936..9155  
/note="MIR repeat: matches 8..241 of consensus"  
9122..9174  
/note="L12 repeat: matches 2668..2735 of consensus"  
9898..9919  
/note="11 copies 2 mer at 100% conserved"  
complement(10194..11028)  
/note="match: GSS: Em:AQ738844"  
10451..10635  
/note="MER5A repeat: matches 2..184 of consensus"  
11814..12073  
/note="L12 repeat: matches 2468..2750 of consensus"  
12588..12796  
/note="MIR repeat: matches 23..236 of consensus"  
15726..16149  
/note="L1MB3 repeat: matches 5050..5473 of consensus"  
16261..16628  
/note="M1T1A1 repeat: matches 1..363 of consensus"  
17825..18145  
/note="Aluub repeat: matches 1..311 of consensus"  
complement(18234..18628)  
/note="match: GSS: Em:AQ549175"  
18512..18622  
/note="MER2 repeat: matches 6..113 of consensus"

```

repeat_region 18843..18929
  /note="MER5A repeat: matches 91..171 of consensus"
repeat_region 20014..20350
  /note="THE1B repeat: matches 1..364 of consensus"
repeat_region 20477..20532
  /note="7 copies 8 mer cacacacc 80% conserved"
repeat_region 20490..20531
  /note="21 copies 2 mer ac 83% conserved"
repeat_region 20501..20532
  /note="8 copies 4 mer caca 93% conserved"
repeat_region 21385..21805
  /note="LMC4 repeat: matches 7288..7768 of consensus"
repeat_region 22623..22714
  /note="MIR repeat: matches 39..131 of consensus"
repeat_region 22800..22931
  /note="LIME3A repeat: matches 6015..6157 of consensus"
repeat_region 23151..23198
  /note="24 copies 2 mer ga 72% conserved"
repeat_region 23918..23978
  /note="MIR repeat: matches 28..90 of consensus"
repeat_region 24098..24217
  /note="15 copies 8 mer ttctctt 63% conserved"
repeat_region 24152..24215
  /note="16 copies 4 mer tcct 68% conserved"
repeat_region 24365..24469
  /note="MER5B repeat: matches 66..178 of consensus"
repeat_region 25225..25542
  /note="AlusX repeat: matches 1..307 of consensus"
repeat_region 25746..26044
  /note="AluY repeat: matches 1..307 of consensus"
repeat_region 26883..26885
  /note="clone RP11-500011
gac in this entry
substitution"
repeat_region 26988..27196
  /note="LMC4 repeat: matches 6626..6848 of consensus"
repeat_region 27203..27350
  /note="MIR repeat: matches 237..389 of consensus"
repeat_region 27478..27557
  /note="10 copies 8 mer aaaggag 67% conserved"
repeat_region 27479..27535
  /note="19 copies 3 mer gag 71% conserved"
repeat_region 27672..27674
  /note="clone RP11-500011
tga in this entry
substitution"
repeat_region 28086..28195
  /note="L2 repeat: matches 2623..2736 of consensus"
repeat_region 28412..28608
  /note="MIR repeat: matches 40..250 of consensus"
repeat_region 29144..29255
  /note="Charlie2 repeat: matches 1..114 of consensus"
repeat_region 29304..29378
  /note="Charlie2 repeat: matches 233..306 of consensus"
repeat_region 29664..29666
  /note="clone RP11-500011
aac in this entry
substitution"
variation 29784..29786
  /replace="agc"
variation 29849..29851
  /replace="cgt"
variation 29849..29851
  /note="clone RP11-500011
acg in this entry
substitution"
repeat_region 30483..30544
  /note="31 copies 2 mer tt 72% conserved"

```

```

variation 30629..30631
  /note="clone RP11-500011
aac in this entry
substitution"
repeat_region 30668..30795
  /replace="agc"
variation 32436..32438
  /note="MIR repeat: matches 41..173 of consensus"
variation 33016..33018
  /replace="cca"
variation 33016..33018
  /note="clone RP11-500011
cct in this entry
substitution"
/replace="ctt"

```

```

Query Match 6.3%; Score 20; DB 9; Length 117757;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 47 aaatgaatcgtttctcca 66
Db 37386 AATGAATCAGTTTCCTCA 37367

```

Search completed: September 16, 2002, 06:31:07  
 Job time: 12345 sec

---

**THIS PAGE BLANK (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 03:05:22 : Search time 3813.96 Seconds  
(without alignments)  
9217.868 Million cell updates/sec

Title: US-09-700-770-2

Perfect score: 1680

Sequence: 1 ggtcgtcagagataaagtt.....tcaataaacacttgctctgt 1680

Scoring table: OLIGO\_MNC  
Gapop 60.0, Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl1:  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	1628	96.9	1680	6	AX224643	Sequence
2	1457	86.7	1636	6	AX092346	Sequence
3	1457	86.7	2036	6	AX224647	Sequence
4	1355	80.7	1684	9	BC008429	Homo sapi
5	562	33.5	1035	6	AX301908	Sequence
6	405	24.1	963	6	AX301906	Sequence
7	181	10.8	159440	9	AL355392	Human DNA
8	161	9.6	161593	9	HSBA49G10	Human DNA
9	26	1.5	26	6	AX301961	Sequence
10	26	1.5	89536	2	AC094188	Rattus no
11	24	1.4	1629	10	MMU46068	Mus musculu
12	22	1.3	200000	6	AP000494	Homo sapi
13	21	1.2	21	6	AX301962	Sequence
14	21	1.2	667	10	AF184920	Sequence
15	21	1.2	32369	2	AC096278	Rattus no
16	21	1.2	75382	2	AC022091	Homo sapi
17	21	1.2	114748	2	AC097411	Rattus no
18	21	1.2	150207	2	AL451067	Homo sapi
19	21	1.2	156012	2	AC008453	Homo sapi
20	21	1.2	169045	2	AC092972	Homo sapi
21	21	1.2	178377	9	AC007685	Homo sapi
22	21	1.2	187651	2	AC055122	Homo sapi
23	21	1.2	187727	2	AC021142	Homo sapi
24	21	1.2	188527	9	AC084357	Homo sapi
25	21	1.2	231743	2	AC095186	Rattus no
26	21	1.2	248087	2	AC097742	Rattus no
27	20	1.2	500	9	IR057017	Homo sapi
28	20	1.2	1321	1	AF246363	Acidithio
29	20	1.2	1698	10	AB041660	Mus muscu
30	20	1.2	1788	9	HSHP512	Human PHS1-
31	20	1.2	1961	6	AB060916	Macaca fa
32	20	1.2	8032	6	AX069334	Sequence
33	20	1.2	12827	2	AC107283	Rattus no
34	20	1.2	20058	2	AC023166	Mus muscu
35	20	1.2	59617	8	AC007087	Arabidops
36	20	1.2	71225	9	AL358196	Human DNA
37	20	1.2	71476	2	AP000490	Homo sapi
38	20	1.2	72921	2	AC102273	Mus muscu
39	20	1.2	140596	2	RN75P15	Rattus no
40	20	1.2	144333	2	AC097962	Rattus no
41	20	1.2	147109	9	AL591043	Human DNA
42	20	1.2	149394	2	AP000828	Homo sapi
43	20	1.2	150000	2	AC095238	Rattus no
44	20	1.2	162481	2	AC091864	Homo sapi
45	20	1.2	171304	2	AC090297	Homo sapi

## ALIGNMENTS

RESULT 1  
LOCUS AX224643 1680 bp DNA linear PAT 10-SEP-2001  
DEFINITION Sequence 1 from Patent WO0161055.  
ACCESSION AX224643  
VERSION AX224643.1 GI:15554772  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 1680)  
AUTHORS Chen, S. Y., Sun, Y. and Macina, R. A.  
TITLE Methods for diagnosing, monitoring, staging, imaging and treating lung cancer via lung cancer specific genes  
JOURNAL Patent: WO 0161055-A 1 23-AUG-2001;  
diabexus, Inc. (US)  
FEATURES  
SOURCE location/Qualifiers  
1. 1680  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT 380 a 503 c 429 g 367 t 1 others  
ORIGIN

Query Match 96.9%; Score 1628; DB 6; Length 1680;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1628; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 53 gggagcgccgaggaactccagcgtgcccaggttcgcatccctcgtcgtcctcgtcga 112
DB 53 GGAAGCGGCGGAGACTCCAGCGCTGCCAGGCTGCGACCTCTGCTGCTGCTGCTGCA 112
QY 113 caccctggaaagatggccgcccgtggaactccctcctcctcgttctgttctgagcga 172
DB 113 CACCTGGGAAGATGGCCGCGCCGCTGAGCTTACCTCTCTGCTGCTGCTGCTGCA 172
QY 173 ccttgatccaagcaccctcagtcaccatgctcctcctcctcgttctgagcga 172
DB 173 CCTTGATCCAAAGCCACCCCTCACTCCAGTGTCTATCTCTGCGCCCAAAATCATCA 232
QY 233 aagaaagctgacacaggaactgaaagacacacagcaccagatccttgaagcgtgc 292
DB 233 AAGAAAGCTGACACAGAGCTGAAGGACACAAAGCCACACAGCATCTCCAGAGCTGC 292
QY 293 cgtctcagtgccatgctggggaagccagcggaggaatccctgtgtgagcagcgtgc 352
DB 293 CGCTCTCAGTGCATGCGGGAAGAACGCGGAGAGATCCCTGTGCTGGGCAAGCTTGGT 352
QY 353 gaacacgclccatgaagcaccatcatcgtgaaagtatcaacagctaaacatccctcagct 412
DB 353 GAACACCGTCTGAAAGCATCATCTGCTGAAGGTCTATCAGACTAACATCTCTCCAGCT 412
QY 413 gcaagtgaaagccctcggcccaatgacccagagctgctagtcgaatccctcgtgagct 472
DB 413 GCAAGTGAAGCCCTCGGCAATGACCGAGAGTGTAGTCAATGATCCCTCTGGACATGCT 472
QY 473 ggtctgatacaacacgcccctgtgtaagaccatcgtgaggttccacatgaacagctgagc 532
DB 473 GGCTGATTCACACCGCCCTGTGTCAAGACCATCTGTGAGTTCCACATGACAGATGAGGC 532
QY 533 ccaagccacacatccgcatgagacacagctgcaagtggcccccagcgcctgttctcagta 592
DB 533 CCAAGCCACCATCCGATGAGACACAGCTCAAGTGGCCCCACCCCTGTGCTCTAGTGA 592
QY 593 ctgtgcacacagcactggaagcctggacatccaaactgctgtaagctcctcctcgt 652
DB 593 CTGTGCACACAGCATGGAGACCTGTGCAATCTGTGATGAAGCTCTCTCTCTGCT 652
QY 653 gaaagccttaagtaagcaagatgataacatcctcctcgtcctcgtcctcgtcgtcgt 712
DB 653 GAAAGCCTTAAGCTAAGAGGTATGACCTCTCTGATGATGATGATGATGATGATGATG 712
QY 713 aaacagcctgtgtccctgcatcggagctcctcctcctcctcctcctcctcctcctcctc 772
DB 713 AAACAGCTGTGTCCCTGATGAGGCTTCTTCAATGAGCATGTATGATGATGATGATGATG 772
QY 773 gctgtgtaaggtgcccattccctcctcctcctcctcctcctcctcctcctcctcctcct 832
DB 773 GCTGTGAAGGTGCCATTTCCTCAGCATTTGACCGCTCTGATGATGATGATGATGATGATG 832
QY 833 tgcacatcaaggtgacacacatcctcctcctcctcctcctcctcctcctcctcctcctcct 892
DB 833 TGCCATCAAGGCTGACACCATCAAGCTTACGTTGGGGGCCAATGTTGATGATGATGATG 892
QY 893 aaaggtgacaaagtgtgtcaataactctgagcttccctcctcctcctcctcctcctcct 952
DB 893 AAAGGTGACCAAGTGTGTAATTAACCTGTGAGCTTCCCTGACATCCACACCTTGACAA 952
QY 953 catcccgcttaccctatcgttgaagcaagctggtgaaagtgtgagctggtcgtcgtcgt 1012
DB 953 CATCCGCTTACCGCTATGCTGAGTCAAGAGCTGTGGAAGCTGACAGTGGCTGCTGCT 1012
QY 1013 ctctcagaagaatcatgtcctgtgtgagctgtgtcctcctcctcctcctcctcctcct 1072
DB 1013 CTCTCCAGAAGATTCATGCTGTGTGACTCTGTGCTCTCTGAGAGTGGCCATCGGCT 1072

```

```

QY 1073 gaagtcagacatgggctgataatgaagaagctgcagaataagctggatctaccagat 1132
DB 1073 GAAGTCAGACATGGGCTGATCATATGAAGAGCTGAGTAAAGCTGAGATCTACCCAGAT 1132
QY 1133 cgtgaagatcctaactcaagacacatcccgagtttcttaagacaaagcactgcagaagt 1192
DB 1133 CGTGAAGATCTCACTCAAGACACCTCCGAGTTTATAGCCAAAGGCCATGCCAAGGT 1192
QY 1193 ggcaccaatgctgctggaagtgttcctcctcctcctcctcctcctcctcctcctcctc 1252
DB 1193 GGCACCAATGATCGTGTGGAAGTTCCTCCAGTGAAGCTCCGCTTGTGTCTCAG 1252
QY 1253 cctggagcctgagaagcactcggaaagctcagttttaaaccaaagtgtacaaattact 1312
DB 1253 CCTGGGCAATCGAAGCCAGCTGGAAAGCTCACTTTTACACCAAGGTGACCACTTATAC 1312
QY 1313 caacttgaataacatcaactctgatacagctgatacagctgatacctcgtgagctggtt 1372
DB 1313 CAACTTGAATTAACATCACTCTGATCGGATCCAGCTGATGAACCTGTGGATTGCTGTT 1372
QY 1373 ccaacatgatttctgaaaaaacaatcactcactgagatcactcactcactcgtgcgaa 1432
DB 1373 CCAACGATATGTTCTGAAAAACATCATCACTGAGATCATCACTCATCTCTGCTGCGAA 1432
QY 1433 ccagaatggaacaaatlaagatctgggtgccagtgatcattgtggaagccttggattga 1492
DB 1433 CCAGAAATGGCAAAATTAAGATCTGGGGTCCAGTGTATTTGTAAGGCTTGGATTGCA 1492
QY 1493 ggcagcagatcctcactgacccaagatgacctgtgttactcagcctccttggaa 1552
DB 1493 GGCAGCTGATCTCTCACTGACCAAGAGATGCCCTTGTGCTTACTCCAGCCTCTTGGA 1552
QY 1553 acccagctcctcgttcccaagtgaaagcttgatggagcagcactaaggaagctggctc 1612
DB 1553 ACCAGCTCTCCTGCTCTCCAGTGAAGATTTGATGCGACAGCATGAGGAGGCTGGGTC 1612
QY 1613 ccagcctggagatagtggtgtgagcctctatagacacatccctcctcctcctcctcct 1672
DB 1613 CCAGCTGGAGATGAGGTGTGAGCTTATAGCCATWCCCTCTGTGCAATATTAACACT 1672
QY 1673 tgctgtg 1680
DB 1673 TGCTGTG 1680

```

RESULT 2  
 AX092346  
 LOCUS  
 DEFINITION  
 AX092346  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE  
 1 (bases 1 to 1636)  
 Eaton,D.L., Flivaroff,E., Gerritsen,M.E., Goddard,A.,  
 Godowski,P.J., Grimaldi,C.J., Gurney,A.L., Watanabe,C.K. and  
 Wood,W.I.  
 TITILE  
 Secreted and transmembrane polypeptides and nucleic acids encoding  
 the same  
 JOURNAL  
 Patent: WO 0116318-A 77 08-MAR-2001;  
 Genentech, Inc. (US)  
 FEATURES  
 source  
 1..1636  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 BASE COUNT 374 a 492 c 411 g 359 t  
 ORIGIN



Query Match		86.7%;	Score 1457;	DB 6;	Length 1636;
Best Local Similarity		99.9%;	Pred. No. 0;		
Matches 1627;		Conservative	0;	Mismatches	1; Indels 1; Gaps 1;
Qy	53	ggaagcgagcgaaggaatccagcgtgcccaggtctgcatctctgcatctgctgctccttga	112		
Db	3	GGAGGGGCGCGAGCATCCAGCGTGCAGGTCGTGCATCTGACACTTGTCTCCCTCTGA	62		
Qy	113	caaccgggaagatggtgcgcccgtgagaccttcaacctctctgtgtgtgtgtgtgcaagca	172		
Db	63	CACCTGGGAAGATGGCCGCGCCGCTGGACCTTCCCTCTCTGTGTGTGTGTGTGTGTGTGT	111		
Qy	173	ccttataccaagccacccctcagctccactgagttccatctctctgagcccaaaagtcacca	232		
Db	123	CCTTATATCCAAAGCCACCTCACTGCTCCACTGCACTTCTATCTCTCGGCCCAAAAGTCATCA	182		
Qy	233	aagaaagctgaacagagagctgaaagaaacaaacagccacagcaatctctgacagcagctgc	292		
Db	183	AAGAAAGCTGACACAGAGAGCTGAAGAGACACAAAGCCACAGCATCTGACAGCAGCTGC	242		
Qy	293	cgctgtcagtgccatctggtggaagaaagccagccgagag -atccctgtgtgtgtgtgtgtgt	351		
Db	243	CGTGTCTAGTGTCTGT	302		
Qy	352	tgaacaccgtctcctggaagacatcatctgtgtgaaagtcatcaacagctaacatctccagc	411		
Db	303	TGAACACCGTCTCTGAAGACATCATCTGTGTGAAGTATCATACAGCTTATCTCTCCAGC	362		
Qy	412	tgcaggtgaagccctcggccaatgaaacagagagctgtgtagtaagaatcccccctgagacatg	471		
Db	363	TGCAGGTGAAGCCCTCGGCGCAATGACAGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGT	422		
Qy	472	tgcgtgagttcaaaagccctctgtgtcaagacacatctgtgtgtgtgtgtgtgtgtgtgtgtgt	531		
Db	423	TGGCTGGATTAACACGCGCCCTGTGTCAAGACCATGTGTGAAGTTCACATGAAGACTGTGAG	482		
Qy	532	cccaagcaccatccgcgaatgaaacaaagtgaagtgagccaccccgctgtgtgtgtgtgtgtgt	591		
Db	483	CCCAAGCACCATCCGAGTGGACACACATGCAAGTGGCCCAACCGCGCTGTGTGTGTGTGTGT	542		
Qy	592	actgtgccaagcaatgagagctgtgcatcaacatgctgtgcatatgaatctctctctgtgt	651		
Db	543	ACTGTGCCACAGCAGCATGGAGCCGTGCGCATTCACACTGCTGTATTAAGCTCTCTCTCTCGG	602		
Qy	652	tgaagcctttagctaaaggttcaatgaacccctctagttgcaatccctctgcccacatctagtga	711		
Db	603	TGAAGCCTTAGCTTAAGGATGAGTCAATGAACCTCTAGTCCATCCCTGCCAATCTAATGA	662		
Qy	712	aaacacagctgtgtccgtgtgacagagcttccctcaatcagcagctgcatgcaagccctctgtc	771		
Db	663	AAACACAGCTGTGTCCGT	722		
Qy	772	agctgtgtgaaggtgccatcttccctcagcatctgacagctgtgagtttgaacctctgtatc	831		
Db	723	AGCTGTGTGAAGGTGCCCTTTCCCTCACCATTTGACCTGTGAGTTGACCTCTGTATTC	782		
Qy	832	cttgcatataaaggtgagacacatctgagctctacccctgtgagggccaaagtgtgtgtgaatcaag	891		
Db	783	CTGCCATTAAGGGTGACACCATTCAGCTCTACCTGTGGGGCCAAAGTGTGTGTGTGTGTGTGT	842		
Qy	892	gaaaggtgacaaagtgttcaataactctgacagcttccctgtgacaaatgcccaccccgagca	951		
Db	843	GAAAGGTGACCAAGTGTTCATAATCTGTGACGCTTCCCTGACATATGCCACCTCGAGCA	902		
Qy	952	acatccgttcaagctcatctgt	1011		
Db	903	ACATCCCTTCAAGCTCTCATCTGT	962		
Qy	1012	tctctcagaagaatctgttctctgt	1071		
Db	963	TCTCTCCGAAGAATTCATGT	1022		
Qy	1072	tgaagtcagaagatcggt	1131		

Db	1023	TCAACTCAAGCATTCGGCGTGCATCATTAAGAAAGCTGCGAGTATAGCTGGATTTACCCAGA	1082		
Qy	1132	tcgtgaagatccctaactaagcaactccaggttttttataagacaagggcatgccaaag	1191		
Db	1083	TCGTGAAGATCCTTAACACAGACACTCCCGAGTTTATTAAGACAAAGCCATTCGCAAG	1142		
Qy	1192	tggcccaactgtctgt	1251		
Db	1143	TGGCCCACTGTATCTGT	1202		
Qy	1252	ccctgtgcatcgaagccagctcggaagctcaagtttataccaaagtgacaactatac	1311		
Db	1203	CCCTGGGATTCGAACCGACCTGGAAGCTCACTTTTACACCAAGTGAACCACTTATAC	1262		
Qy	1312	tcaacttgaataacalcagctctgtgcatccagctgagatgaaactctggaatgtgtgt	1371		
Db	1263	TCAACTGAATTAACATCAAGCTCTGTATCGGATCCAGCTGATGAACCTGTGGATTGGCTGGT	1322		
Qy	1372	tccaacctgagtcttctgaaaaacatcatcactgagatcatcactccactctgtgtgcga	1431		
Db	1323	TCCAACCTGATGTCTGTGAAGAAACATCATCATGAGATCATTCACATCATCTGTGTGTGTGT	1382		
Qy	1432	accagaatggaatcaatgaatctgt	1491		
Db	1383	ACCAGATGTGCAATTAAGATCTGGGCTCCAGTGTCAATGTGTGAAGGCTTGGGATTCG	1442		
Qy	1492	aggcagctgagctcctcactgacgacaaagatgcccctgtgtgtgtgtgtgtgtgtgtgtgtgt	1551		
Db	1443	AGGCAGCTGACGCTCTACATGACCAAGATGCTCTGTGTGTGTGTGTGTGTGTGTGTGTGT	1502		
Qy	1552	aaccagctctctgtcttcccgagtaagactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1611		
Db	1503	AAOCCAGCTCTCTGTCTCCAGTGAAGACTTGAGTGGAGCATCAAGGAAGGCTGGGT	1562		
Qy	1612	cccaagctggaatgaggt	1671		
Db	1563	CCCACTGGGATTAAGGAGTGTGAGCTATATACCATCTCTGTGTGTGTGTGTGTGTGTGTGT	1622		
Qy	1672	ttgctgt			
Db	1623	TTGCTGT			

RESULT 3					
AX224647	LOCUS	AX224647	2036 bp	DNA	linear
AX224647	DEFINITION	Sequence 5 from Patent WO0161055.			
AX224647	ACCESSION	AX224647			
AX224647.1	VERSION	GI:15554774			
KEYWORDS					
SOURCE		human.			
ORGANISM		Homo sapiens			
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
AUTHORS		Chen, S.Y., Sun, Y. and Macina, R.A.			
TITLE		Methods for diagnosing, monitoring, staging, imaging and treating Lung Cancer via Lung cancer specific genes			
JOURNAL		Patent: WO 0161055-A 5 23-AUG-2001;			
FEATURES		diadexus, Inc. (US)			
SOURCE		Location/Qualifiers			
		1. 2036			
		/organism="Homo sapiens"			
		/db_xref="taxon:9606"			
BASE COUNT		477 a	586 c	528 g	444 t
ORIGIN		1 others			
Query Match		86.7%;	Score 1457;	DB 6;	Length 2036;
Best Local Similarity		99.9%;	Pred. No. 0;		
Matches 1627;		Conservative	0;	Mismatches	1; Indels 1; Gaps 1;

QY	53	ggaagggccgaaggaactccgaagctgcccaggtctcgaagctctgcaactctgtctgcacctga	112
Db	92	ggagcgggcccaggaactccgaagctggcccaagctctggcattccatctgctgacctctgga	151
QY	113	caacttggaaatgtgcccgcaccgttgaaacctcaacctctctctgtgtgttgctgcaacca	172
Db	152	CACTGGGAAATGGCCGGCCCGGGACCTTCACTCTCTCTGTGTGTGGTGGCAACCA	211
QY	173	ccttgatccaaagccaccttaagtcccaatgtccaatgtccatctcctcctggcccacaaatca	232
Db	212	CCTTGATCCAAAGCCACCTTATGTCCACATGCAAGTTCTCATCTCTGGGCCAAATGATCA	271
QY	233	.aagaanaacttgaacaaagagcttgaagacacaaacgcacacagatctctgpaagatgtc	292
Db	272	AAGAAAGCTGACACAGAGACTGAAGACACCAACAGCCACAGATCCCTGAGAGCTGG	331
QY	293	gcgtgcctcagttgcacatgctgcgggaaagccagccggagag-atccctgtctgtggcagctg	351
Db	332	CGCTGCTAGTGCATGTGCGGGAAACCCAGCCGGAGGAGATCCCTGTCTGGCAGCCTGG	391
QY	352	tgaacaccgcctcgtgaagacacatcatctgtgcctgaagatcaacagctcaaatctccagc	411
Db	392	TGAACACCGTCTGTGAAGCACATCATCTGGCTGAAGATCATACAGCTAAATCTTCCAC	451
QY	412	tgcaggtgtgaagccctctgcgcacatgacscasgagctgtctaagtaaatcccccttgaacatg	471
Db	452	TGCAAGGTGAAGCCCTTGGCCAAATGACACAGAGAGCTGTCAATGAATGCCCTTGACATGG	511
QY	472	tggctgatttcaacaacgcgccctcgtgtcgaagacatctgtgagttccaatgaagactgaag	531
Db	512	TGGCTGATTCAAACAGCGCCCTGTGTCAAGACCATCTGTGGATTCACATGACACATGAG	571
QY	532	cccaagcccaacctcgcgatgtgaacacacagttgaagtggccccaccgcgcgtgtccctgaag	591
Db	572	CCCAACCCACCAATCCGATGAGACACACAGTGAATGTGGCCCCACCGCCTGTGCTCCAGG	631
QY	592	acgtgtccacacagccatggagagccctgcgcacatccaactcgtcgtcaataagctctcctctg	651
Db	632	ACGTGTCCACCAAGCCATGGAGACCTTGGCATCAACGTGCAATGAAGCTCTCTTCCCTGG	691
QY	652	tgaagccttgaagcaagctgtacacgtgtcaactctcctagtccatccctgcgccaatctaga	711
Db	692	TGAAGCCCTTACTGTAAGCAGGTCTATGAACCTCCGTAAGGCATCCCGCCCAATGTAGTGA	751
QY	712	aaaacagagctgtgtccggctgtgttgaaagcttccctcaatgaatgcatgtatgcaagactctgc	771
Db	752	AAAACAGAGTGTGTCCCGTGTGAGGAGCTTCTTCAATGACATGATGAGACACCTCTGC	811
QY	772	agctgtgtgaaggtgtgccatttccctcctcgaatgacagctctgagtttgaactctcgtatc	831
Db	812	AGCTGTGTGAAGGTGCCATTTCTCCCTCAGATGACGCTGTGAAGTTTGACTTCTGATTC	871
QY	832	ctgcacatgaagggtgtgaacacatcagctcactctactctggtgggcccagttgttgaactoaag	891
Db	872	CTGCATGAAGGTTGACACATTCATGCTCACTCTGGGGCCCAAGTGTGTGAGCTACACAG	931
QY	892	gaaaggtgaaccaagtgttcaataaactctgcagcttccctcctgaatgacatgcccacacctgaca	951
Db	932	GAAAGGTGACCAAGTGTTCATTAACCTGCAAGCTTCCCTGACATGCCACCTTGACA	991
QY	952	acatcccgcttaagctcatatgttgaafcaagaagctgtgtgaagactgcagctgtcgtctgtc	1011
Db	992	ACATCCCGCTTACGCTCATGTATGTGAATGAGAGAGCTGATGAACCTGCAATGTGCTCTGTGC	1051
QY	1012	tctctcagaagaatttcatgtctcgttctgtgactctgtctcctctcgaagatggtcccatgtgc	1071
Db	1052	TCTCTCCAGAAATTCATGTGCTCTTGTGACTCTGTGCTTCTGTAGAGTGCCTCATGGC	1111
QY	1072	tgaagttaagcatcgggctgtgataatlgaaagagctgcagataaagcttggatctcaccaga	1131
Db	1112	TGAAGTGAACATCGGAGCTGATCAATGAAGAAAGGCTGCAGATTAACGTGGATTAACCCAGA	1171
QY	1132	tcgtgaagatccttaactcgaagcaactcccgagttttttatatagaccaaaggccatgtccaag	1191

Db	1172	TCGTGAAGATCTCTACCTACGAGCACTCCCGAGTTTATATAGACCAAGGCCATGCGAAGG	1231
QY	1192	tggtcccaactgatcgtgctgtgaaagtgtttccctccagtgaaagccctccgaccttggta	1251
Db	1232	TGGCCCAACTATCTGCTGGAAAGTGTTCCTCCAGTGAAGACCCTCGCCCTTGTTC	1291
QY	1252	cctctggatcgaagccagcctcgcgaagctcagttttacacaaagatgacaaacttaac	1311
Db	1292	CCCTGGGCGATCGAAGCCAGCGCTCGGAGCTCAGTTTACACCAAGAGTACCAACTTATAC	1351
QY	1312	tcaacttgataatacaatcagctcgtatcgcagatcgcagatgaactctggatctgctgt	1371
Db	1352	TCAACTTGAATTAACATCAGCTGCTCATGGAGTCCAGCTGATGATCACTCTGGATGGCTGTGT	1411
QY	1372	tccaacttgatctctgaaaaaacatcatcatatgataatcatccatccatctctgtccga	1431
Db	1412	TCCAACCTGATGTTCTGAAAAACATCATCTGATGATATCATCACTCCATCTCGTGGCGA	1471
QY	1432	accagaatgcaaatataagatctgggggtccagtgatcatgtgtgtaagagcttggatctg	1491
Db	1472	ACCAAGATGCGAAATTAAGATCTGGGGTCCACAGTGTATGTGGTAAGCCCTTGGGATTCG	1531
QY	1492	aggagatctgagctctcaactacaaagaatcccttggcttactccaagctcctgttga	1551
Db	1532	AGGAGCTGATGCTCTCACTACCAAGGATCCCTTGCTGTACTCCAGCCCTCTTGGA	1591
QY	1552	aaccagctctcctgtctcccaatgaaagcttgatgagcgcatacaggaagagctgggt	1611
Db	1592	AACCCACACTCTCCTGTCTCCACTGAAGACTTGATGGAGGCATCAGGAAAGCTGTGGT	1651
QY	1612	ccagagctgggaatgagtggtgtgagctctataagccatccctctctgcaatcaataaac	1671
Db	1652	CCAGCTGGAGATGAGGTGTGAGCTCTATAGACCATCCCTCTCAATCAATAAACAC	1711
QY	1672	ttagcctgtg 1680	
Db	1712	TTGCTGTGTG 1720	
RESULT	4		
LOCUS	BC008429	1684 bp	mRNA linear PRI 12-JUL-2001
DEFINITION		Human sapiens, similar to DNA segment, Chr 2, Messerschmitts Institute of Technology 19, clone MGC:14597 IMAGE:4291561, mRNA, complete cds.	
ACCESSION	BC008429		
VERSION	BC008429.1	GI:14250057	
KEYWORDS	MGC.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 1684)		
TITLE	Strausberg R		
JOURNAL	Direct Submission		
COMMENT	Submitted (25-MAY-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: <a href="http://mgc.ncl.nih.gov">http://mgc.ncl.nih.gov</a>		
COMMENT	Contact: MGC help desk Email: <a href="mailto:cgabs-remail.nih.gov">cgabs-remail.nih.gov</a> Tissue Procurement: CLONTECH cDNA Library Preparation: CLONTECH Laboratories, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINI) DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305 Web site: <a href="http://www.shgc.stanford.edu">http://www.shgc.stanford.edu</a> Contact: (Dickson, Mark) <a href="mailto:mdcpaxil.stanford.edu">mdcpaxil.stanford.edu</a> Dickson, M., Schmitt, J., Grimwood, J., Rodriguez, A., and Myers, R. M.		



```

QY 1672 ttgctctgtg 1680
      |||||
Db 1626 ttgctctgtg 1634

RESULT 5
LOCUS AX301908 1035 bp DNA Linear PAT 30-NOV-2001
DEFINITION Sequence 15 from Patent WO0174851.
ACCESSION AX301908
VERSION AX301908.1 GI:17382958
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Majumder,K., Spaderna,S.K., Taupier,R.J., Padigar,M.,
Burgess,C.E., Shinkets,R.A., Spytek,K.A., Liu,X., Patturajan,M. and
Gusev,V.Y.
TITLE Novel proteins and nucleic acids encoding same
JOURNAL Patent: WO 0174851-A 15 11-OCT-2001;
Curagen Corporation (US)
FEATURES
source 1..1035
location/Qualifiers
CDS
/organism="Homo sapiens"
/db_xref="taxon:9606"
79..1035
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD13090.1"
/db_xref="GI:17382958"
/translation="MAGPWFTLLCGLLAATLIQATLSPTAVLILGPKYIKETLQEL
KDNHATSILOQLPLSAMREKPAIGIPVLSIVNTYLIKHIIMKVTANILQIOLVPS
ANDDELIVKIPIDMWAGFNTPIVKTIIEFMTAEATITMDTSAGPTRIIVSDCAT
SHGSLRIOLHLKLSFLVNAIAKOVNMLVPSLIVKNQOLCPVTEASFGNVAIDLQI
VKGRCASLSPFSTFTEILASRPGKVTKMFNNSASLTMPITDNIIPFSLIVSODVYKAA
VAAVAVLSPPEEFVLLDSVNLSTRORIGPFRPHRNFLNTGCP"
BASE COUNT 230 a 333 c 261 g 211 t
ORIGIN

Query Match 33.5%; Score 562; DB 6; Length 1035;
Best Local Similarity 99.9%; Pred. No. 1e-303;
Matches 682; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 102 ctgcccttgacacctggygaagatgcccggccgtggaaccttcaacctctctgtgtt 161
      |||||
Db 57 ctgcccttgacacctggaagatggccggccgtggaaccttcaacctctctgtgtt 116

QY 162 gctggagcaaccttgatcaagcaacctcagtcacacagtggttctatctctggccc 221
      |||||
Db 117 gctggagcaaccttgatcaagcaacctcagtcacacagtggttctatctctggccc 176

QY 222 aaaaatcatcaaaagaagctgacacagagctgaaggaacacacagcaccagcatct 281
      |||||
Db 177 aaaaatcatcaaaagaagctgacacagagctgaaggaacacacagcaccagcatct 236

QY 282 gcaagagctggccgtctgacagtcacagtcgaggaagaaagccagccgagag-attccctgtact 340
      |||||
Db 237 gcaagagctggccgtctgacagtcacagtcgaggaagaaagccagccgagagctctgtgtact 296

QY 341 gggcagcctgtgtgaacacgctcctgaagacacatctgtcgtgaagtgatatacagctaa 400
      |||||
Db 297 gggcagcctgtgtgaacacgctcctgaagacacatctgtcgtgaagtgatatacagctaa 356

QY 401 catctccagctgagagtgtaagccctcgagcaatagcaagagctgtactaagatccc 460
      |||||
Db 357 catctccagctgagagtgtaagccctcgagcaatagcaagagctgtactaagatccc 416

QY 461 ccttgacatgtgtgtgtatcaacacgcccctgtgtcaagacacatcgtgaggttccacat 520
      |||||
Db 417 ccttgacatgtgtgtgtatcaacacgcccctgtgtcaagacacatcgtgaggttccacat 476

```

```

QY 521 gacagctgaagggcccaagccacacatccgcatgagacacagctgaagtgagcccaagccct 580
      |||||
Db 477 gacagctgaagggcccaagccacacatccgcatgagacacagctgaagtgagcccaagccct 536

QY 581 ggtcctcagtgactgtgtccacacagcaatggagagctgtgcataccactgtctgataagct 640
      |||||
Db 537 ggtcctcagtgactgtgtccacacagcaatggagagctgtgcataccactgtctgataagct 596

QY 641 ctccctcgtgtgaagcccttaagcaggtgcatgaactctctgtgcaatccctgac 700
      |||||
Db 597 ctccctcgtgtgaagcccttaagcaggtgcatgaactctctgtgcaatccctgac 656

QY 701 caatctagtgaaaaaaacagctgtgtcccgatgacagagctctcctcaatgagcatgac 760
      |||||
Db 657 caatctagtgaaaaaaacagctgtgtcccgatgacagagctctcctcaatgagcatgac 716

QY 761 agaacctctcagctgtgtgaag 783
      |||||
Db 717 agaacctctcagctgtgtgaag 739

RESULT 6
LOCUS AX301906 963 bp DNA Linear PAT 30-NOV-2001
DEFINITION Sequence 13 from Patent WO0174851.
ACCESSION AX301906
VERSION AX301906.1 GI:17382956
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Majumder,K., Spaderna,S.K., Taupier,R.J., Padigar,M.,
Burgess,C.E., Shinkets,R.A., Spytek,K.A., Liu,X., Patturajan,M. and
Gusev,V.Y.
TITLE Novel proteins and nucleic acids encoding same
JOURNAL Patent: WO 0174851-A 13 11-OCT-2001;
Curagen Corporation (US)
FEATURES
source 1..963
location/Qualifiers
CDS
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..963
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD13089.1"
/db_xref="GI:17382957"
/translation="MAGPWFTLLCGLLAATLIQATLSPTAVLILGPKYIKETLQEL
KDNHATSILOQLPLSAMREKPAIGIPVLSIVNTYLIKHIIMKVTANILQIOLVPS
ANDDELIVKIPIDMWAGFNTPIVKTIIEFMTAEATITMDTSAGPTRIIVSDCAT
SHGSLRIOLHLKLSFLVNAIAKOVNMLVPSLIVKNQOLCPVTEASFGNVAIDLQI
VKGRCASLSPFSTFTEILASRPGKVTKMFNNSASLTMPITDNIIPFSLIVSODVYK
AAVAAVAVLSPPEEFVLLDSVNLSTRORIGPFRPHRNFLNTGCP"
BASE COUNT 217 a 314 c 237 g 195 t
ORIGIN

Query Match 24.1%; Score 405; DB 6; Length 963;
Best Local Similarity 100.0%; Pred. No. 1.7e-215;
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 379 ggtcgaaggtatcaacagctaaacatccctcagcgtgcaagtggaagccctggccaatgacc 438
      |||||
Db 263 ggtcgaaggtatcaacagctaaacatccctcagcgtgcaagtggaagccctggccaatgacc 322

QY 439 agagagctgtagtaagaatccctcgtgacatgtgtgtgtatcaacagccctgtgta 498
      |||||
Db 323 agagagctgtagtaagaatccctcgtgacatgtgtgtgtatcaacagccctgtgta 382

QY 499 agaacatcgtgtgaggttccacatgagcactgaagcccaagccacatccgcatgagacca 558
      |||||

```

Db	383	AGACCATGTGGAGTTCACATGAGACTGAGAGCCCAAGCACCATTCGCATGAGACCA	442
Qy	539	gtgcaagtggcccaaccgctgtgctcctaagtactgtgcaaccagcactggagctgtc	618
Db	443	GTGCAGATGGGCCCCACCGGCTGGTCTTCACAGTGCACAGCCATGGAGACCTTGC	502
Qy	619	gcattcaactctgtgataagctctctcttcctggtgaaagccttagctaaagcgtc	678
Db	503	GCATCAACTGCTGTATAGCTCTCTTCCTGGTGAACGCTTACTTAAGCAGGTCATGA	562
Qy	679	acctctagtcccaatccctgtcccaatctagtgaanaacacagctgtccctgtacga	738
Db	563	ACCTCTAGTGTGCATCCCTGCCCATCTAGTGAATAAACACAGCTGTCTCCGTCGAT	622
Qy	739	cttctcctaattgacatgataatgacagacccctcctgacgtgtgtaagg	783
Db	623	CTTCTCTCAATGGCATGTATGACAGACCTCTGTCGACCTGTGAAG	667
RESULT	7		
LOCUS	AL355392	159440 bp	DNA linear PRI 09-MAR-2001
DEFINITION	AL355392	Human DNA sequence from clone RP5-118774 on chromosome 20q11.1-11.23 Contains ESTs, STSS, GSSS and two Cpg islands. Contains the gene for novel protein similar to mouse von Ebner salivary gland protein, the gene for a novel protein similar to rat RYF3, the LOC51654 gene for a novel protein (GCI-05) similar to rat CDK5 activator-binding protein and the SMT1 gene for alpha syntrophin (dystrophin-associated protein A1, 59Kd, acidic component), complete sequence.	
ACCESSION	AL355392.7	GI:10178502	
VERSION	AL355392.7	GI:10178502	
KEYWORDS	HTG; GCI-05; Cpg Island; RYF3; SMT1; SMTA1; syntrophin.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	1 (bases 1 to 159440)		
TITLE	Bird.C.		
JOURNAL	Direct Submission		
COMMENT	Submitted (06-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk On May 14, 2001 this sequence version replaced gi:6982468 gi:9213549. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the validation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at http://www.sanger.ac.uk/RGP/Chr20 This sequence is the entire insert of clone RP5-118774 The true left end of clone RP11-327D19 is at 62286 in this sequence. The true right end of clone RP11-49G310 is at 100 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP5-118774 is from the library RPL-5 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm		

```

FEATURES
source
VECTOR: pcyrpac2
Location/Qualifiers
1..159440
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="q11.1-11.23"
/clone="RP5-1187J4"
/clone_id="RPC1-5"
1535..1848
/note="157 copies 2 mer ga 61% conserved"
1757..1824
/note="Tandem repeat. weak data"
join(3408..3471,5782..5867,7116..7295,8242..8295,
8799..8957,9841..9908,10729..10774,11885..11948,
12794..12870,15591..15761)
/gene="dj1187J4.1"
/note="match: ESTs: Em:BE933124 Em:BE061388"
/product="dj1187J4.1.1 (novel protein similar to mouse von
Ebner salivary gland protein, isoform 1.)"
/evidence=not_experimental
3408..15761
/gene="dj1187J4.1"
join(<3408..3471,5782..5867,7116..7295,8242..8295,
8799..8957,9841..9908,10729..10774,11885..11948,
12794..12870,15591..15650)
/gene="dj1187J4.1"
/note="Continues in Em:AL121901 as ba49G10.6"
/codon_start=1
/evidence=not_experimental
/product="dj1187J4.1.1 (novel protein similar to mouse von
Ebner salivary gland protein, isoform 1.)"
/protein_id="CAC34050.1"
/db_xref="GI:13274680"
/translation="ILCPVIEASFNGVADLLQLYKVPISISIDRLFEEDLYLPKIGDT
IQLYLAKGLIDSGCKVYKRNNSASLTMTLTIINPSLIYSQDYAAVAAYSPSE
PMVLDSVLPESARLKSSIGLINERADKLGQIYKIIITQDTPPEFIDQGHAKNAQ
LIVLEVPSSSEALRPFTLTGIEASSGEQFTYKGPDLILNLNNISSDRIOLMNSGIGWF
QPDVLKNIITEIHISILLPQNQKLRSGVSVYKALGFEEAESSLTAKDLVYTPASTL
WRPSPYSQ"
3817..3915
/misc_feature
3817..3915
/gene="dj1187J4.1"
/note="match: STS: Em:G10242"
/complement(join(3917..4006,31119..31326))
/note="match: STS: Em:U75536"
3917..4143
/misc_feature
3917..4143
/gene="dj1187J4.1"
/note="match: STS: Em:G10242"
4662..4691
/misc_feature
4662..4691
/note="15 copies 2 mer tt 86% conserved"
join(5758..5867,7116..7295,8242..8295,8799..8957,
9841..9908,10729..10774,11885..11948,12794..12870,
15591..15761)
/misc_feature
15591..15761
/gene="dj1187J4.1"
/note="match: cDNAs: Em:U46068
match: ESTs: Em:AI924632 Em:AW363590 Em:AW384404
Em:AW449208 Em:AW167650 Em:AW381927 Em:AW167610
Em:AW577643 Em:AW176323 Em:AI024451 Em:U45555 Em:AW176329
Em:AW176327 Em:AW176325 Em:AI909965 Em:AI909964
Em:AA428580 Em:AW384405"
/product="dj1187J4.1.2 (novel protein similar to mouse von
Ebner salivary gland protein, isoform 2.)"
/evidence=not_experimental
/complement(6761..7253)
/note="match: GSS: Em:AQ149447"
join(7182..7295,8242..8295,8799..8957,9841..9908,
10729..10774,11885..11948,12794..12870,15591..15650)
/gene="dj1187J4.1"
/codon_start=1
/evidence=not_experimental
/product="dj1187J4.1.2 (novel protein similar to mouse von
Ebner salivary gland protein, isoform 2.)"
/protein_id="CAC15880.1"

```

```

/db_xref="GI:11137741"
/translation="MPTLDNIPFSLIVSQDVKAANAVALSPSEFVLLDSVLPSAH
RLKSSIGLINERKADKLRSQIVKILIODPEPFIDGHAKVADLIYLEPSSERAR
PLFLTGIFASAEQGYRKGDOLIINLNNISSDRLOLNNSGTGPVPLKNIITETIH
SILPLNONGKTRSGVPSIVKALFEAEASSLTIDALVLPASTLMKPSFVSQ"
10970..11009
repeat_region
/feature="20 copies 2 mer ac 85% conserved"
complement(11466..11766)
misc_feature
/feature="match: GSS: Em:AQ060606"
complement(11631..11769)
misc_feature
/feature="match: GSS: Em:AQ427662"
11766..12192
misc_feature
/feature="dJ1187J4.1"
11768..12108
misc_feature
/feature="match: GSS: Em:AQ403818"
11775..12108
misc_feature
/feature="match: GSS: Em:AQ373301"
11775..12108
misc_feature
/feature="dJ1187J4.1"
14322..14833
repeat_region
/feature="256 copies 2 mer ca 68% conserved"
14836..14871
repeat_region
/feature="18 copies 2 mer ac 88% conserved"
complement(15031..15459)
misc_feature
/feature="match: GSS: Em:AQ694930"
15736..15741
polyA_signal
/feature="dJ1187J4.1"
15754
polyA_site
/feature="dJ1187J4.1"
15761
polyA_site
/feature="dJ1187J4.1"
15761
misc_feature
/feature="match: GSS: Em:AQ349609"
complement(17617..18004)
misc_feature
/feature="match: GSS: Em:AQ350711"
18109..18565
misc_feature
/feature="match: GSS: Em:HS81207T"
complement(20661..21112)
misc_feature
/feature="match: GSS: Em:AQ019396"
complement(20690..21115)
misc_feature
/feature="match: GSS: Em:AQ178024"
21105..21680
misc_feature
/feature="match: GSS: Em:HS156A22T"
complement(22577..22968)
misc_feature
/feature="match: GSS: Em:AQ508829"
22977..23573
misc_feature
/feature="match: GSS: Em:AQ281488"
complement(37780..38128)
misc_feature
/feature="match: GSS: Em:HSF13F11"
complement(39112..39718)
misc_feature
/feature="match: GSS: Em:AQ636647"
complement(39266..39748)
misc_feature
/feature="match: GSS: Em:AQ827010"
39747..40042
misc_feature
/feature="match: GSS: Em:AQ282137"
39754..40132
misc_feature
/feature="match: GSS: Em:AQ252720"
39763..40037
misc_feature
/feature="match: GSS: Em:AQ029439"
match: STS: Em:G50113"
complement(39950..40165)
misc_feature
/feature="match: GSS: Em:AL207848"
complement(39950..40334)
misc_feature
/feature="match: GSS: Em:AQ556676"
39961..40217
misc_feature
/feature="match: GSS: Em:AQ143633"
complement(39964..40169)
misc_feature
/feature="match: STS: Em:G02503"
complement(39991..40409)
misc_feature
/feature="match: GSS: Em:AQ596761"
complement(40002..40172)
misc_feature
/feature="match: GSS: Em:AL185783"

```

```

misc_feature      complement(40006..40317)
/feature="match: GSS: Em:AQ052600"

Query Match
Best Local Similarity 100.0%; Score 181; DB 9; Length 159440;
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 869 ggcgaattgttggtacccacagggaaagtgtgaccagtgttaataacttcagcttc 928
DB 7115 GGCCTAATTGTTGGACACACAGGGAAGGTGACCAAGTGTCAATTAACCTCAGACTTC 7174
QY 929 cctgacatgcccacccctgtgacacacatccctgtcagctcactatgtgaagcagctgt 988
DB 7175 CCGACAAATGCCCCACCTGTGACACATCCCTTCAGCTCATGTGATGAGCAGCGTGGT 7234
QY 989 gaaagctgcaagtgtgtctgtctctccagaagaattcatgtctctgtgacctgt 1048
DB 7235 GAAAGCTGCAGTGGCTGCTGTCTCTCCAGAAATTCATGCTGCTGTGACACTGTG 7294
QY 1049 g 1049
DB 7295 G 7295

```

```

RESULT 8
HSBA49G10
LOCUS
DEFINITION
HSBA49G10 161593 bp DNA linear PRI 20-JUN-2001
Human DNA sequence from clone RP11-49G10 on chromosome 20. Contains
a gene similar to bovine salivary protein BSP30, the LOC51297 gene
for LUNX protein, 5' end of a gene encoding a protein similar to
murine von ebner minor salivary gland protein, a novel gene, a
pseudogene similar to STAT-induced STAT inhibitor-2, a pseudogene
similar to ribosomal protein l12, a putative novel transcript,
ESTs, STSs, GSSs and a Cpg island, complete sequence.
AL121901.20 GI:8249854
HTG: BSP30; Cpg island; LOC51297; LUNX; RP112; STAT-inhibitor.
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 161593)
REFERENCE
AUTHORS
Tracey A.
TITLE
Direct Submission
JOURNAL
Submitted (14-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Jun 5, 2000 this sequence version replaced gi:8017404.

```

COMMENT

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr20>

IMPORTANT: This sequence is not the entire insert of clone RP11-49G10. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true left end of clone RP11-49G10 is at 1 in this sequence. The true left end of clone RP5-1187J4 is at 161494 in this sequence. The true right end of clone RP4-7330Z3 is at 27823 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing

problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. Rpl1-49G10 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6.  
Location/Qualifiers

```

FEATURES
    source
        1..161593
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="20"
        /clone="Rpl1-49G10"
        /clone_lib="RPCI-11.1"
        complement(1..107)
        /note="match: GSS: Em: A0607627"
        6..706
        /note="match: GSS: Em: A0051854"
        137..548
        /note="match: GSS: Em: A0550113"
        530..603
        repeat_region
            /note="37 copies 2 mer ca 66% conserved"
            533..686
            /note="2 copies 77 mer 83% conserved"
            613..678
            repeat_region
                /note="33 copies 2 mer ca 75% conserved"
                3030..3174
                /note="l1p83 repeat: matches 5941..6084 of consensus"
                5140..5183
                repeat_region
                    /note="22 copies 2 mer tt 75% conserved"
                    5184..5344
                    /note="l1pA2 repeat: matches 5986..6146 of consensus"
                    complement(6104..6463)
                    /note="match: GSS: Em: A0339300"
                    6120..6239
                    /note="2 copies 60 mer 85% conserved"
                    complement(8476..9002)
                    /note="match: GSS: Em: A0537190"
                    complement(14766..15061)
                    /note="match: STS: Em: H8922114S"
                    14772..14779
                    /note="Random repeat. Forced join. Approximately 500 bases missing according to restriction digest."
                    16219..16387
                    misc_feature
                        /note="match: GSS: Em: AL001671"
                        16225..16387
                    misc_feature
                        /note="match: GSS: Em: AL013146"
                        18325..18826
                    misc_feature
                        /note="match: GSS: Em: A0458133"
                        18328..18828
                    misc_feature
                        /note="match: GSS: Em: A0532957"
                        18333..18770
                    misc_feature
                        /note="match: GSS: Em: A0765930"
                        complement(20975..21245)
                    misc_feature
                        /note="match: GSS: Em: A2102442"
                        20999..21238
                    repeat_region
                        /note="4 copies 60 mer 69% conserved"
                        21031..21263
                    misc_feature
                        /note="match: GSS: Em: A2055779 Em: A2084220"
                        21144..21225
                    repeat_region
                        /note="41 copies 2 mer ag 75% conserved"
                        complement(21172..21428)
                        /note="match: GSS: Em: A0395327"
                        complement(23473..23950)
                    misc_feature
                        /note="match: GSS: Em: A0727877"
                        complement(23637..23948)
                    misc_feature
                        /note="match: GSS: Em: A0691469"
                        23950..24355
                    misc_feature
                        /note="match: GSS: Em: A0152593"
                        23955..24456
                    misc_feature
                        /note="match: GSS: Em: A0508832"
                        26979..27422
                    misc_feature
                        /note="match: GSS: Em: A0636646"

```

```

repeat_region
    31882..32061
    /note="3 copies 60 mer 73% conserved"
repeat_region
    31921..32032
    /note="56 copies 2 mer ct 63% conserved"
repeat_region
    33693..33762
    /note="MER47 repeat: matches 2251..2322 of consensus"
    33733..33785
    /note="MER47 repeat: matches 1..55 of consensus"
    join(35535..35585,36507..36678,40308..40452,41455..41562,42783..42935,45523..45604,46980..47043,47896..47936)
    48600..48787
    /gene="DA49G10.1"
    /note="match: CDNAS: Em: U79414 Em: U79413 Em: AX061621 match: ESTs: Em: A1654622"
    /product="DA49G10.1 (similar to bovine salivary protein Bsp30)"
    /evidence="not_experimental"
    35335..48787
    /gene="DA49G10.1"
    join(36522..36678,40308..40452,41455..41562,42783..42935,45523..45604,46980..47043,47896..47936)
    /gene="DA49G10.1"
    /note="match: proteins: Tr:P79125 Sw:P07743"
    /codon_start=1
    /evidence="not_experimental"
    /product="DA49G10.1 (similar to bovine salivary protein Bsp30)"
    /protein_id="CAC03546.1"
    /db_xref="GI:9801234"
    /translation="MLQLMKLVLLGCVLTGTSSESLDNLGNDLSNVVDKLPVLAHEGL ETVDDTLKGLIEKLVLDGVLOKSSAPMAOKAEAKLNNVSKLLPNTDIFGL KINSLLIDVKAEPIDOGKGNISPPVAVNTVAGPIGOIINLKASLDLTAATIER DPQHQPVAVIGECASDPTSLSLDHSQINKFVNSVINTLKSYSLLQKEIC LIRFIHSLDVAVNIQVVDNPDHKTQDTLT"
    38155..38751
    /note="l1pA8 repeat: matches 4514..5125 of consensus"
    38750..39785
    /note="l1pA8 repeat: matches 5128..6163 of consensus"
    48769..48774
    /gene="DA49G10.1"
    48787
    /gene="DA49G10.1"
    52540..53450
    /note="Cpg island"
    /evidence="not_experimental"
    complement(53851..54131)
    /note="match: GSS: Em: A0897745"
    53831..54141
    /note="match: STS: Em: G31731"
    53842..53969
    /note="match: GSS: Em: A0429347"
    complement(53844..54141)
    /note="match: GSS: Em: AF101963"
    53858..54022
    /note="match: GSS: Em: A0392891"
    complement(53908..54141)
    /note="match: GSS: Em: A0415287"
    53958..54139
    /note="match: GSS: Em: A0378081"
    53970..54139
    /note="match: GSS: Em: A0527291"
    53972..54142
    /note="match: GSS: Em: A0429347"
    54000..54141
    /note="match: GSS: Em: A0392510"
    complement(54027..54139)
    /note="match: GSS: Em: A0533658"
    54030..54138
    /note="match: GSS: Em: AG000935"
    54041..54149
    /note="match: GSS: Em: A0392891"
    54767..55668
    /gene="DA49G10.2"

```





\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
4714 4713: contig of 4713 bp in length  
4814 4813: gap of unknown length  
7443 7442: contig of 2629 bp in length  
7543 7542: gap of unknown length  
8971 8970: contig of 1428 bp in length  
9071 9070: gap of unknown length  
11618 11618: contig of 2548 bp in length  
11719 11718: gap of unknown length  
14064 14064: contig of 2346 bp in length  
14164 14164: gap of unknown length  
14165 14165: contig of 2434 bp in length  
16599 16598: gap of unknown length  
16599 16598: gap of unknown length  
16599 16598: contig of 1860 bp in length  
18659 18658: gap of unknown length  
21636 21636: contig of 2978 bp in length  
21736 21736: gap of unknown length  
23091 23091: contig of 1355 bp in length  
23191 23191: gap of unknown length  
24363 24363: contig of 1172 bp in length  
24463 24463: gap of unknown length  
26616 26616: contig of 2153 bp in length  
26716 26716: gap of unknown length  
29328 29328: contig of 2612 bp in length  
29428 29428: gap of unknown length  
30861 30861: contig of 1433 bp in length  
30961 30961: gap of unknown length  
32666 32666: contig of 1705 bp in length  
32766 32766: gap of unknown length  
34751 34751: contig of 1985 bp in length  
34851 34851: gap of unknown length  
36722 36722: contig of 1871 bp in length  
36822 36822: gap of unknown length  
37990 37990: contig of 1168 bp in length  
38090 38090: gap of unknown length  
38951 38951: contig of 1761 bp in length  
39852 39852: gap of unknown length  
41482 41482: contig of 1531 bp in length  
41582 41582: gap of unknown length  
42970 42970: contig of 1388 bp in length  
43071 43071: gap of unknown length  
44717 44717: contig of 1647 bp in length  
44817 44817: gap of unknown length  
46731 46731: contig of 1914 bp in length  
46831 46831: gap of unknown length  
48721 48721: contig of 1890 bp in length  
48821 48821: gap of unknown length  
50484 50484: contig of 1663 bp in length  
50584 50584: gap of unknown length  
52333 52333: contig of 1749 bp in length  
52433 52433: gap of unknown length  
53718 53718: contig of 1285 bp in length  
53818 53818: gap of unknown length  
55081 55081: contig of 1263 bp in length  
55181 55181: gap of unknown length  
55437 55437: contig of 1256 bp in length  
56438 56438: gap of unknown length  
56537 56537: gap of unknown length  
58395 58395: contig of 1858 bp in length  
58495 58495: gap of unknown length  
60176 60176: contig of 1681 bp in length  
60276 60276: gap of unknown length  
62186 62186: contig of 1910 bp in length  
62286 62286: gap of unknown length  
64498 64498: contig of 2212 bp in length  
64598 64598: gap of unknown length  
65989 65989: contig of 1391 bp in length  
66089 66089: gap of unknown length  
67493 67493: contig of 1404 bp in length  
67593 67593: gap of unknown length  
69641 69641: contig of 2048 bp in length

\* 69642 69741: gap of unknown length  
\* 69742 70862: contig of 1121 bp in length  
\* 70863 70962: gap of unknown length  
\* 70963 72389: contig of 1427 bp in length  
\* 72390 72489: gap of unknown length  
\* 72490 74082: contig of 1593 bp in length  
\* 74083 74182: gap of unknown length  
\* 74183 75218: contig of 1036 bp in length  
\* 75219 75318: gap of unknown length  
\* 75319 76809: contig of 1491 bp in length  
\* 76810 76909: gap of unknown length  
\* 76910 78708: contig of 1799 bp in length  
\* 78709 78808: gap of unknown length  
\* 78809 79849: contig of 1041 bp in length  
\* 79850 79949: gap of unknown length  
\* 79950 81538: contig of 1589 bp in length  
\* 81539 82748: gap of unknown length  
\* 82749 82848: contig of 1110 bp in length  
\* 82849 84058: gap of unknown length  
\* 84059 84158: gap of unknown length  
\* 84159 85695: contig of 1537 bp in length  
\* 85696 85795: gap of unknown length  
\* 85796 86875: contig of 1080 bp in length  
\* 86876 86975: gap of unknown length  
\* 86976 88063: contig of 1088 bp in length  
\* 88064 88163: gap of unknown length  
\* 88164 89536: contig of 1373 bp in length.

Location/Qualifiers  
1. .89536  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/clone="CH230-2K16"

BASE COUNT 21918 a 20067 c 20357 g 22306 t 4888 others

ORIGIN

Query Match 1.5%: Score 26; DB 2; Length 89536;  
Best Local Similarity 100.0%; Pred. No. 0.012;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 456 atccccctgacatgctgctgalt 481  
|||||  
Db 23604 ATCCCCCTGCAGCATGTGCTGCATT 23579

RESULT 11  
MMU46068 1629 bp mRNA ROD 10-AUG-2000  
LOCUS  
DEFINITION Mus musculus von Ebner minor salivary gland protein mRNA, complete cds.  
ACCESSION U46068  
VERSION U46068.2 GI:9789706  
KEYWORDS  
SOURCE house mouse strain-Swiss-Webster.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 1629)  
AUTHORS Sneed,M.L., Villanueva,J., Paine,M.L., Lei,Y.P., Zhu,D.H.,  
Luis,J., Xia,Y.-R. and Yang,J.-N.  
TITLE Direct Submission  
JOURNML Submitted (17-JAN-1996) Malcolm L. Sneed, Center for Cranio  
Molecular Biology, University of Southern California, 2250 Alcazar  
St, Los Angeles, CA 90033, USA  
COMMENT On Aug 10, 2000 this sequence version replaced gi:1184789.  
FEATURES  
source  
1. .1629  
/organism="Mus musculus"  
/strain="Swiss-Webster"  
/db\_xref="taxon:10090"  
/chromosome="2"  
/map="2; D2Mit19n and D2Mit25n"

CDS

58..1482  
 /codon\_start=1  
 /product="von Ebner minor salivary gland protein"  
 /protein\_id="AAA87581.2"  
 /db\_xref="GI:9789707"

translation="MAGPWITLLCGLLGATLVQANVYPAVLNLGPEVIOKHLTOAL  
 KNDHTAIIQDELPLRAMODKSGIPILDSSEVAVLKIIIMKVTSAIIQLDVQPSR  
 YNOELVRIPLDMVAGNLPLIKITVEFMSTEVQALIRVRSKSGPAHMLSDCSN  
 ESTRLSLHKLSFVNSLAKNVMNLVLPALPOLVKNLCPMOADVDWYEDPLRT  
 TAPIALSPGALFGLSPAIODSNILMLKKILDSARVYNNNSATSIEMETTPG  
 APFLVYKQDLYNALVTWLVPRKEVILLRPVITDVAKQLOMDIKETNAEKANLIGPT  
 OMLKFTSTPHIVLNESGARAASVILEVPTNDVRFESLGIASYSKQPTEDN  
 RLMMFNNSVIERIKMLSDILKLDPEVLKDTLRLKILEYLLPNNCKLRTGIPMSMS  
 KALGTERKMSVSKGALNLTTPASS"

BASE COUNT 411 a 497 c 382 g 339 t

ORIGIN

Query Match 1.4%; Score 24; DB 10; Length 1629;  
 Best Local Similarity 100.0%; Pred. No. 0.19;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 456 atccccctgacatgctgctgga 479  
 |||||  
 Db 388 ATCCCCCTGACATGCTGCTGGA 411

RESULT 12  
 AP000494  
 LOCUS AP000494 200000 bp DNA linear PRI 28-SEP-1999  
 DEFINITION Homo sapiens genomic DNA, chromosome 3p21.3, clone:301 to 308,  
 anti-oncogene region, section 2/5.  
 ACCESSION AP000494  
 VERSION AP000494.1 GI:5926671  
 KEYWORDS  
 SOURCE Homo sapiens DNA.  
 ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (sites)  
 Nakamura,Y., Isomura,M., Daigo,Y., Tamari,M. and Ishikawa,S.  
 DNA sequence analysis of a 1.2-Mb region on chromosome 3p21.3  
 2 (bases 1 to 200000)  
 Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.  
 Direct Submission  
 Submitted (21-SEP-1999) to the DDBJ/EMBL/GenBank databases. Miki  
 Hirakawa, Japan Science and Technology Corporation (JST), Advanced  
 Databases Department, 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081,  
 Japan (E-mail:mikae@tokyo.jst.go.jp,  
 URL:http://www-alls.tokyo.jst.go.jp/, Tel:81-3-5214-8470,  
 Fax:81-3-5214-8470)  
 This sequence is conducted by Japanese Foundation for Cancer  
 Research as a JST sequencing team.  
 Principal Investigator: Yusuke Nakamura Ph.D  
 Phone:81-3-5449-5372, Fax:81-3-5449-5433,  
 yusuke@nci.ims.u-tokyo.ac.jp  
 The sequence is submitted by Human Genome Sequencing in ALLS  
 project of JST  
 Japan Science and Technology Corporation (JST)  
 5-3, Yonbancho, Chiyoda-ku, Tokyo, 102-0081 Japan  
 For further information about this sequences, please visit our  
 sequence archive Web site (http://www-alls.tokyo.jst.go.jp/HGS/top.  
 html) or send email to webmaster@www-alls.tokyo.jst.go.jp.  
 Location/Qualifiers  
 1..200000  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="3"  
 /map="3p21.3"  
 /note="301-308"  
 complement(112666..112810)  
 /note="sts69202:The location is between each flanking site

## FEATURES

Source

STS

of PCR primers."  
 /db\_xref="GDB:4569933"  
 BASE COUNT 52190 a 42809 c 46061 g 58756 t 184 others  
 ORIGIN

Query Match 1.3%; Score 22; DB 9; Length 200000;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 842 ggggtgacacatcagctctac 863  
 |||||  
 Db 135252 GGGTGACACCATTCAGCTCTAC 135273

RESULT 13  
 AX301962/c  
 LOCUS AX301962 21 bp DNA linear PAT 30-NOV-2001  
 DEFINITION Sequence 69 from Patent W00174851.  
 ACCESSION AX301962  
 VERSION AX301962.1 GI:17382977  
 KEYWORDS  
 SOURCE synthetic construct.  
 ORGANISM synthetic construct.  
 artificial sequence.  
 1 (sites)  
 REFERENCE  
 AUTHORS Majumder,K., Spaderna,S.K., Taupier,R.J., Padigar,M.,  
 Burgess,C.E., Shmukets,R.A., Spytek,K.A., Liu,X., Paturajan,M. and  
 Gusev,V.Y.  
 Novel proteins and nucleic acids encoding same  
 Patent: WO 0174851-A 69 11-OCT-2001;  
 Curagen Corporation (US)

FEATURES  
 SOURCE location/Qualifiers  
 1..21  
 /organism="synthetic construct"  
 /db\_xref="taxon:32630"  
 /note="Ag719 Reverse Primer Sequence"

BASE COUNT 4 a 6 c 5 g 6 t

Query Match 1.2%; Score 21; DB 6; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 432 aatgacagagcgcgcagtc 452  
 |||||  
 Db 21 AATGACGAGAGCTGCTAGTC 1

RESULT 14  
 AF184920  
 LOCUS AF184920 667 bp mRNA linear ROD 05-OCT-1999  
 DEFINITION Rattus norvegicus GDNF-family receptor alpha 3 (GFRalpha-3) mRNA,  
 partial cds.  
 ACCESSION AF184920  
 VERSION AF184920.1 GI:6010760  
 KEYWORDS  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

REFERENCE 1 (bases 1 to 667)  
 AUTHORS Stover,T., Gong,T.-W., Cho,Y., Altschuler,R.A. and Lomax,M.I.  
 TITLE Expression of neurturin, artemin, persephin and their receptors  
 JOURNAL GFRa-2 and GFRa-3 in the mature rat cochlea  
 REFERENCE 2 (bases 1 to 667)  
 AUTHORS Stover,T., Gong,T.-W., Cho,Y., Altschuler,R.A. and Lomax,M.I.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-SEP-1999) Kresge Hearing Research Institute,  
 University of Michigan, 9301E MSRB III, 1150 West Medical Center

FEATURES  
Location/Qualifiers  
1..667  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/tissue\_type="cochlea; substantia nigra"  
1..>667  
/gene="GfRalpha-3"  
1..>667  
/gene="GfRalpha-3"  
/product="GDNF-family receptor alpha 3"  
/protein\_id="AAFO1242.1"  
/db\_xref="gi:6010761"  
/translation="TVTSKPKMKMLSKSLMKPDSGLCKFAMLTLDRCDRLEKAY  
GEACGICQRIHCLAQLRFFEKAEASHAGCLLCPAPEDAGGERRRNTIASCA  
LPVAPNCLDLSRFRADPLRCRLMDQTHCHPMDILGTCATBOSRCLRAYLIGT  
AMTPFISKVTVALGCTCRSGNLODECEQLKESQFNCLMEALIAKMRFHRLF  
SODW"

BASE COUNT 138 a 222 c 175 g 132 t

ORIGIN

Query Match 1.2%; Score 21; DB 10; Length 667;  
Best Local Similarity 100.0%; Pred. No. 9.4;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 tggacttcagaccacatgcc 40  
|||||  
Db 384 TGGACTTCCAGACCCACTGCC 404

RESULT 15  
AC096278 32369 bp DNA linear HTG 20-DEC-2001  
DEFINITION Rattus norvegicus clone CH230-11L24, \*\*\* SEQUENCING IN PROGRESS  
ACCESSION AC096278.2 GI:17943975  
VERSION AC096278.2  
KEYWORDS HTG, HTGS, PHASE1.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 32369)  
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbara,J.,  
Benton,J., Bimagne,K., Blankenburg,K., Bonnin,D., Bouck,J.,  
Bowen,S., Briteva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,  
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,  
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,  
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,  
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
Degen,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,  
Dugan-Hocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,  
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,  
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,  
Gasta,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,  
Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A.,  
Hernandez,J., Hernandez,O., Hodgson,A., Hughes,M., Holloway,C.,  
Hollins,B., Homsi,F., Howard,S., Huber,J., Huix,S., Hume,J.,  
Jachson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jollivel,S.,  
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,  
Kovar,C., Kralovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,  
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,  
Loulsegh,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,  
Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,  
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,  
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,

Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,  
Ogurb,M., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,  
Peery,J., Perez,L., Peters,L., Pickens,R., Prins,E., Pu,L.L.,  
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojudoen,I., Rolle,M.,  
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshbari,N.,  
Sisson,I., Sodergren,E., Sonalka,T., Sparks,A., Stanley,H.,  
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,  
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,  
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,  
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,  
Watlington,S., Williams,G., Williamson,A., Wleczky,R., Wooden,S.,  
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
Weinstock,G. and Gibbs,R.  
Direct Submission  
Unpublished  
2 (bases 1 to 32369)  
Worley,K.C.  
Direct Submission  
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Dec 20, 2001 this sequence version replaced gi:15627898.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GEVR  
Center clone name: CH230-11L24  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329Flrst call to  
findphraplist  
Consensus quality: 25575 bases at least Q40  
Consensus quality: 29387 bases at least Q30  
Consensus quality: 32762 bases at least Q20  
Estimated insert size: 11642; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; basecall-tp estimation  
Quality coverage: 0.1x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 21 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1  
\* 2640: contig of 2640 bp in length  
\* 2641  
\* 2740: gap of unknown length  
\* 2741  
\* 4533: contig of 1793 bp in length  
\* 4534  
\* 4633: gap of unknown length  
\* 4634  
\* 6203: contig of 1570 bp in length  
\* 6204  
\* 6303: gap of unknown length  
\* 6304  
\* 7920: contig of 1617 bp in length  
\* 7921  
\* 8020: gap of unknown length  
\* 8021  
\* 9285: contig of 1265 bp in length  
\* 9286  
\* 9385: gap of unknown length  
\* 10442: contig of 1057 bp in length  
\* 10443  
\* 10542: gap of unknown length  
\* 10543  
\* 12042: contig of 1500 bp in length  
\* 12043  
\* 12142: gap of unknown length  
\* 12143  
\* 13857: contig of 1715 bp in length  
\* 13858  
\* 13957: gap of unknown length  
\* 13958  
\* 15637: contig of 1680 bp in length  
\* 15638  
\* 15737: gap of unknown length  
\* 15738  
\* 16927: contig of 1190 bp in length  
\* 16928  
\* 17027: gap of unknown length  
\* 17028  
\* 19301: contig of 2274 bp in length  
\* 19302  
\* 20634: contig of 1233 bp in length

```

* 20635 20734: gap of unknown length
* 20735 21741: contig of 1007 bp in length
* 21742 21841: gap of unknown length
* 21842 23151: contig of 1310 bp in length
* 23152 23251: gap of unknown length
* 23252 24576: contig of 1325 bp in length
* 24577 24676: gap of unknown length
* 24677 25830: contig of 1154 bp in length
* 25831 25930: gap of unknown length
* 25931 27156: contig of 1227 bp in length
* 27156 27257: gap of unknown length
* 27258 28348: contig of 1091 bp in length
* 28349 28448: gap of unknown length
* 28449 29559: contig of 1111 bp in length
* 29560 29659: gap of unknown length
* 29660 30777: contig of 1118 bp in length
* 30778 30877: gap of unknown length
* 30878 32369: contig of 1492 bp in length.

```

FEATURES  
Source 1. .32369 Location/Qualifiers

/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/clone="CH230-11L24"

BASE COUNT 8252 a 6711 c 7651 g 7723 t 2032 others  
ORIGIN

Query Match 1.2%; Score 21; DB 2; Length 32369;  
Best Local Similarity 100.0%; Pred. No. 8.2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 280 ctgcagcagctgcgcgctc 300  
|||||  
Db 22520 CTGCAGCAGCTGCCGCTCCTC 22540

Search completed: September 16, 2002, 06:28:35  
Job time: 12193 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 04:24:12 : Search time 390.66 seconds  
(without alignments)  
7385.444 Million cell updates/sec

Title: US-09-700-770-2

Perfect score: 1680

Sequence: 1 ggtgtgcagatataagtt.....tcaataaacacttcctgtg 1680

Scoring table: OLIGO\_NUC

Gapop 60.0, Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size: 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: N.Geneseq\_032802.\*

- 1: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT.\*
- 2: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT.\*
- 3: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT.\*
- 4: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT.\*
- 5: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT.\*
- 6: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT.\*
- 7: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT.\*
- 8: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT.\*
- 9: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT.\*
- 10: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT.\*
- 11: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT.\*
- 12: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT.\*
- 13: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT.\*
- 14: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT.\*
- 15: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT.\*
- 16: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT.\*
- 17: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT.\*
- 18: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT.\*
- 19: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT.\*
- 20: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT.\*
- 21: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT.\*
- 22: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT.\*
- 23: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT.\*
- 24: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1680	100.0	1680	21	AA229724	Human Lung specific Nucleotide sequenc
2	1628	96.9	1680	22	AAH77948	Human NTAP CDNA cl
3	1458	86.8	1707	21	AA287210	Human PRO1357 (UNO
4	1457	86.7	1636	21	AAA37057	Human DNA encoding
5	1457	86.7	1636	22	AA546064	Human PRO1357 CDNA
6	1457	86.7	1636	22	AAF92096	Human encoding prote
7	1457	86.7	1636	22	AAF54293	Nucleotide sequenc
8	1457	86.7	2036	22	AAH77950	DNA encoding novel
9	1064	63.3	2121	23	AA572707	

10	562	33.5	1035	22	ABA01989	Human NOV6b coding
11	523	31.1	1177	21	AAF18254	Lung cancer associ
12	522	31.1	1175	21	AA297023	Human secreted pro
13	436	26.0	1172	21	AA297114	Human secreted pro
14	420	25.0	420	21	AAH30554	Human COL6A cancer
15	405	24.1	963	22	ABA01986	Human NOV6a coding
16	271	16.1	465	21	AA274959	Human OREF514
17	229	13.6	232	16	AA221378	Human gene STRF14
18	213	12.7	423	22	AA118484	Human breast cance
19	213	12.7	462	22	AA109612	Human breast cance
20	181	10.8	594	22	ABA61288	Human foetal liver
21	181	10.8	594	22	ABA29112	Probe #7578 for ge
22	181	10.8	594	22	AAK09583	Human brain expres
23	181	10.8	594	22	AAK35477	Human bone marrow
24	181	10.8	594	22	AA117050	Probe #6983 for ge
25	181	10.8	594	22	AA141191	Probe #9877 used t
26	180	10.7	180	22	ABA73788	Human foetal liver
27	180	10.7	180	22	ABA38963	Probe #17429 for g
28	180	10.7	180	22	AAK22236	Human brain expres
29	180	10.7	180	22	AAK48404	Human bone marrow
30	180	10.7	180	22	AA126228	Probe #16161 for g
31	180	10.7	180	22	AA154231	Probe #22917 used
32	161	9.6	583	22	ABA60344	Human foetal liver
33	161	9.6	583	22	ABA28591	Probe #7057 for ge
34	161	9.6	583	22	AAK08624	Human brain expres
35	161	9.6	583	22	AAK34508	Human bone marrow
36	161	9.6	583	22	AA116752	Probe #6685 for ge
37	161	9.6	583	22	AA140226	Probe #8912 used t
38	156	9.3	156	22	ABA72883	Human foetal liver
39	156	9.3	156	22	ABA38468	Probe #16934 for g
40	156	9.3	156	22	AAK21316	Human brain expres
41	156	9.3	156	22	AAK47474	Human bone marrow
42	156	9.3	156	22	AA125947	Probe #15880 for g
43	156	9.3	156	22	AA153305	Probe #21991 used
44	150	8.9	576	21	AA274887	Human OREF442
45	141	8.4	375	23	AA572706	DNA encoding novel

## ALIGNMENTS

RESULT 1  
AA229724 standard; DNA; 1680 BP.  
ID AA229724;  
AC AA229724;  
XX  
XX 27-MAR-2000 (first entry)  
XX  
XX  
DE Human Lung specific gene-1.  
XX  
XX  
XX Lung Specific Gene; LSG; human; diagnostic marker;  
XX prognosticate; Lung cancer; diagnosis; ds.  
XX  
XX Homo sapiens.  
XX  
XX  
XX WO960160-A1.  
XX  
XX 25-NOV-1999.  
XX  
XX  
XX 12-MAY-1999; 99WO-US10344.  
XX  
XX  
XX 21-MAY-1998; 98US-0086212.  
XX  
XX  
XX (DIAD-) DIADEXUS LLC.  
XX  
XX Yang F, Macina RA, Sun Y;  
XX  
XX WPI; 2000-116320/10.  
XX  
XX A new method for diagnosing, monitoring and staging lung cancer  
XX  
XX Example 1; Pages 33-34; 40pp; English.  
XX  
XX

XX The present sequence is a lung specific gene (LSG) from human  
CC clone ID 2798946. The LSG has high level of tissue specificity for lungs  
CC and is overexpressed in cancerous tissues. The sequence serves as a  
CC diagnostic marker for detecting, monitoring, staging and prognosticating  
CC lung cancer. The diagnosis involves comparing levels of LSG in samples  
XX obtained from patient and normal control.

SQ Sequence 1680 BP; 380 A; 503 C; 429 G; 367 T; 1 other;

Query Match 100.0%; Score 1680; DB 21; Length 1680;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1680; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ggtgtcaggatataaagttggaactccagaccactgcccgggaagagatgagacgg 60
Db 1 ggtgtcaggatataaagttggaactccagaccactgcccgggaagagatgagacgg 60
QY 61 ccgaggaactccagcgtgcccaggctctggaatccctgactgtgcccctctgacactgg 120
Db 61 ccgaggaactccagcgtgcccaggctctggaatccctgactgtgcccctctgacactgg 120
QY 121 aagaatgcccggccgctggagacccttaccctctctgtgtgtgttctgtgcaacccttgatc 180
Db 121 aagaatgcccggccgctggagacccttaccctctctgtgtgtgttctgtgcaacccttgatc 180
QY 181 caagccaccctcagtcaccactgcaattctcatcctctcgagcccaaaagtcaaaagaagaag 240
Db 181 caagccaccctcagtcaccactgcaattctcatcctctcgagcccaaaagtcaaaagaagaag 240
QY 241 ctgagaccaggaactgtaagaagaccacaacgaccacagcatctctgcaagacgtcgcctgtc 300
Db 241 ctgagaccaggaactgtaagaagaccacaacgaccacagcatctctgcaagacgtcgcctgtc 300
QY 301 agtgcacatgcygggaaagccagccgagagatccctgtgctgcygacgctgtgtgaacaccg 360
Db 301 agtgcacatgcygggaaagccagccgagagatccctgtgctgcygacgctgtgtgaacaccg 360
QY 361 tecttaagacacatcatctgctgctggaagtcatcaacagcttaaacatccctcagctgcaagtga 420
Db 361 tecttaagacacatcatctgctgctggaagtcatcaacagcttaaacatccctcagctgcaagtga 420
QY 421 agccctcgagccaaatgaccaggaagctgtcagatcaagaatccctctggaactgtgtgat 480
Db 421 agccctcgagccaaatgaccaggaagctgtcagatcaagaatccctctggaactgtgtgat 480
QY 481 tcaaacagccctgtgtgaagaccatcgtgtgaagttccacatgaagactgaagcccaagcca 540
Db 481 tcaaacagccctgtgtgaagaccatcgtgtgaagttccacatgaagactgaagcccaagcca 540
QY 541 ccatccgcatgagacaccagctgcaagtggcccacccgctgtgctctgaagtgaactgtgcca 600
Db 541 ccatccgcatgagacaccagctgcaagtggcccacccgctgtgctctgaagtgaactgtgcca 600
QY 601 ccaagccatggagagcctgctgcatccaactgtctgaataagctctctctctctgtgaagcct 660
Db 601 ccaagccatggagagcctgctgcatccaactgtctgaataagctctctctctctgtgaagcct 660
QY 661 taagctaaagcaggtcatgaacctctctagctgcatccctggcccaatctagtgaaaaaccacg 720
Db 661 taagctaaagcaggtcatgaacctctctagctgcatccctggcccaatctagtgaaaaaccacg 720
QY 721 tgtgtccgtgatacgaagctctctcaatgacatgtatgagaacctctgcaactgtgta 780
Db 721 tgtgtccgtgatacgaagctctctcaatgacatgtatgagaacctctgcaactgtgta 780
QY 781 aggtgccatttccctcagcatgtgacgctctggaagttgaacctctgtatctctgcatca 840
Db 781 aggtgccatttccctcagcatgtgacgctctggaagttgaacctctgtatctctgcatca 840
QY 841 aggtgtacacattcagcttacctgaggccaagttgtgtgactacaaggaaagtga 900
Db 841 aggtgtacacattcagcttacctgaggccaagttgtgtgactacaaggaaagtga 900
```

```
Db 841 aggtgtacacattcagcttacctgaggccaagttgtgtgactacaaggaaagtga 900
QY 901 ccaagttgttcaataactctgagcttccctcgacaatgcccacccctgacaacatccct 960
Db 901 ccaagttgttcaataactctgagcttccctcgacaatgcccacccctgacaacatccct 960
QY 961 tcaagcctatcgtgtgaagacgtgtgtgaagactgtgacgtgtgctgtctctcag 1020
Db 961 tcaagcctatcgtgtgaagacgtgtgtgaagactgtgacgtgtgctgtctctcag 1020
QY 1021 aagaattatgtctctgtgttgaaactctgtcttccctgagagtgccatcgctgaagtc 1080
Db 1021 aagaattatgtctctgtgttgaaactctgtcttccctgagagtgccatcgctgaagtc 1080
QY 1081 gcatcggtctgatacaatgaagacgtcagaataagctggagatcccaacaaatgtgaaga 1140
Db 1081 gcatcggtctgatacaatgaagacgtcagaataagctggagatcccaacaaatgtgaaga 1140
QY 1141 tccctaactcaggaacactcccgagtttttatagaaccaagccatgccaaggtggtcccaac 1200
Db 1141 tccctaactcaggaacactcccgagtttttatagaaccaagccatgccaaggtggtcccaac 1200
QY 1201 tgatcgtgtgtaagtgtttccctcagtgaaagccctccgacctgtgttcaaccttgga 1260
Db 1201 tgatcgtgtgtaagtgtttccctcagtgaaagccctccgacctgtgttcaaccttgga 1260
QY 1261 tcgaaagccagctcggaaagctcaagtttaacacaaaggtgacaaactataactaactga 1320
Db 1261 tcgaaagccagctcggaaagctcaagtttaacacaaaggtgacaaactataactaactga 1320
QY 1321 ataacatcagctcgtgacatcgatccagctgaatgaactctggaatgtgtgttccaaactg 1380
Db 1321 ataacatcagctcgtgacatcgatccagctgaatgaactctggaatgtgtgttccaaactg 1380
QY 1381 atgttctgaaacaatcatcatctactgagatcatccatccatctctgtcgcgcgaaccgaatg 1440
Db 1381 atgttctgaaacaatcatcatctactgagatcatccatccatctctgtcgcgcgaaccgaatg 1440
QY 1441 gcaaatatgaatctgggtcccaagtgatcatctgtgtgaagaccttggaatcgagagcagctg 1500
Db 1441 gcaaatatgaatctgggtcccaagtgatcatctgtgtgaagaccttggaatcgagagcagctg 1500
QY 1501 agtctcactgaccaaaggaatgacctgtgtctactccaaagcctctgtggaaccagact 1560
Db 1501 agtctcactgaccaaaggaatgacctgtgtctactccaaagcctctgtggaaccagact 1560
QY 1561 ctccgttctcccaagtgaaagacttgatgagcagcatccagggaaagcttgggtccagctgg 1620
Db 1561 ctccgttctcccaagtgaaagacttgatgagcagcatccagggaaagcttgggtccagctgg 1620
QY 1621 gagtatgggtgtgaagctctatagaccatccctctctgcaatcaataaacacttgcctgtg 1680
Db 1621 gagtatgggtgtgaagctctatagaccatccctctctgcaatcaataaacacttgcctgtg 1680
```

RESULT 2  
ID AAH77948 standard; DNA; 1680 BP.  
XX  
AC AAH77948;  
DE 13-NOV-2001 (first entry)  
XX  
DE Nucleotide sequence of a human lng103 polypeptide.  
XX  
KW Human; lung cancer specific gene; LSG; lng103; lung cancer; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT 124..1577  
FT /\*tag= a  
FT /product= "lng103"

```
FT      " (pos: 328..329, aa: Gly) "  
/transl_except=  
FT      "{pos: 372..374, aa: Val}"
```

```

/translate_except= "(pos: 372..374, aa: Val)"

```

PN W0200161055-A2.

PD 23-AUG-2001.

PF 20-FEB-2001; 2001WO-US05674.

PR 17-FEB-2000; 2000US-0183188.

PA (DIAD-) DIADEXUS INC.

PI Chen S, Sun Y, Macina RA;

DR WPI; 2001-529917/58.

DR P-PSDB; AAG63976.

PT New lung cancer specific gene for the treatment and diagnosis of lung

PT cancer -

PS Claim 1; Page 112-113; 119pp; English.

The present sequence represents a human lung cancer specific gene (LSG), and encodes a polypeptide designated lng103. LSGs are useful in the treatment and diagnosis of lung cancer. The treatment of lung cancer comprises the administration of a molecule which down regulates the expression of an LSG. An immune response can be mounted against a target cell expressing an LSG. Identification of potential therapeutic agents for use in imaging and treating lung cancer which comprises screening molecules for an ability to bind to or decrease expression of an LSG relative to LSG in the absence of the agent where the ability of a molecule to bind to the LSG or decrease expression of the LSG is indicative of the molecule being useful in imaging and treating lung cancer.

SQ Sequence 1680 BP; 380 A; 503 C; 429 G; 367 T; 1 other:

Query Match 96.98; Score 1628; DB 22; Length 1680;

QY	53	ggaagcgggcccgaagatctcaacgtgcccagctctgcatctctgcaactgctgcctctgaa	112
Db	53	ggaagcgggcccgaagatctcaacgtgcccagctctgcatctctgcaactgctgcctctgaa	112
QY	113	caacttggaagaatgcccggcccggtggaacttcaacctctctctgtgtttgtctggcagcca	172
Db	113	caacttggaagaatgcccggcccggtggaacttcaacctctctctgtgtttgtctggcagcca	172
QY	173	ccttgatccaagcccaacctctcaagctcccaatcgatctctatcctctggcccaaaagtcatca	232
Db	173	ccttgatccaagcccaacctctcaagctcccaatcgatctctatcctctggcccaaaagtcatca	232
QY	233	aagaaagaagctgacacacaggaagctgaaagacccaacaacgccaacagcatalctctgcagcgtgc	292
Db	233	aagaaagaagctgacacacaggaagctgaaagacccaacaacgccaacagcatalctctgcagcgtgc	292
QY	293	cgctgcctcaatgacatgcgcggggaaaaagccagcccggaagatccctgtgtcgggcagcctggt	352
Db	293	cgctgcctcaatgacatgcgcggggaaaaagccagcccggaagatccctgtgtcgggcagcctggt	352
QY	353	gaacaacgctctctgaagacacatcatctgctctgaagttcaatcaagttacaatcctccaagt	412
Db	353	gaacaacgctctctgaagacacatcatctgctctgaagttcaatcaagttacaatcctccaagt	412
QY	413	gcacagtgaagccctcgcgacaaatgacccaagagctgtatgaagatcccccctggacatggt	472
Db	413	gcacagtgaagccctcgcgacaaatgacccaagagctgtatgaagatcccccctggacatggt	472
QY	473	ggtctgattcaacaacgcccctggtcaaaacatcgttgatgttccacatgacagcactgagcg	532

Db	473	ggttgattcaacaaagcccttvgttcaagaacatcgttgagtttccacatgaaagactggaagc	532
QY	533	ccaagccacccatccgcatactgagacacagctgcaagctgagccccaacccgccttgctccatgta	592
Db	533	ccaagccacccatccgcatactgagacacagctgcaagctgagccccaacccgccttgctccatgta	592
QY	593	ctgtgtccacacagccatctggagagcctctgcgcatccaactgtctgtcataagctctctctctgt	652
Db	593	ctgtgtccacacacagccatctggagagcctctgcgcatccaactgtctgtcataagctctctctctgt	652
QY	653	gaagccttaagttaagcagagttcataaacctctctagtccatctccgcgcgaactctagttaa	712
Db	653	gaagccttaagttaagcagagttcataaacctctctagtccatctccgcgcgaactctagttaa	712
QY	713	aaacacagctgtgtcccgctgatacgaagctctccctcaatactgscatgtactgagacactctca	772
Db	713	aaacacagctgtgtcccgctgatacgaagctctccctcaatactgscatgtactgagacactctca	772
QY	773	gctgtgtgaagctgtcccaattccctccagaattgagccgctcgtgaaatttgaaactctgtatcc	832
Db	773	gctgtgtgaagctgtcccaattccctccagaattgagccgctcgtgaaatttgaaactctgtatcc	832
QY	833	tgcatctaaaggttgacacacatccatccagctctactctgggggccaagttgttggatacaagg	892
Db	833	tgcatctaaaggttgacacacatccatccagctctactctgggggccaagttgttggatacaagg	892
QY	893	aaaagctgacccaagtggtttcaataactctgcagctccctctgaaatgagccaacccctgaaaa	952
Db	893	aaaagctgacccaagtggtttcaataactctgcagctccctctgaaatgagccaacccctgaaaa	952
QY	953	catcccgcttccagcccatatcgttgagtcagcaagagctgtgtgaaagctgtgacgtgtgtctgt	1012
Db	953	catcccgcttccagcccatatcgttgagtcagcaagagctgtgtgaaagctgtgacgtgtgtctgt	1012
QY	1013	ctctccagaagaatctcatgtgtctgttvgaaactctgtgtctctccagagtgcccatcggct	1072
Db	1013	ctctccagaagaatctcatgtgtctgttvgaaactctgtgtctctccagagtgcccatcggct	1072
QY	1073	gaagtcagaagcatcgggcgtgataatgaaagaagctgtcgaataagctggatactacaacaaat	1132
Db	1073	gaagtcagaagcatcgggcgtgataatgaaagaagctgtcgaataagctggatactacaacaaat	1132
QY	1133	cgtagaagatccctaactcagacatcccgagctttttatagaacaaaggcgaatgcgaagct	1192
Db	1133	cgtagaagatccctaactcagacatcccgagctttttatagaacaaaggcgaatgcgaagct	1192
QY	1193	ggcccaactgatactgtctgtgaaagctgttccctccagtgaaagccctccgctctgttcaac	1252
Db	1193	ggcccaactgatactgtctgtgaaagctgttccctccagtgaaagccctccgctctgttcaac	1252
QY	1253	cctgtgcatctgaaagccagctcggaaagctcagttttacaaccaagaagtgcaccaattatct	1312
Db	1253	cctgtgcatctgaaagccagctcggaaagctcagttttacaaccaagaagtgcaccaattatct	1312
QY	1313	caacttgataaacaacagctctcgtcgcgataccagctgataaactctggattgtgctggtt	1372
Db	1313	caacttgataaacaacagctctcgtcgcgataccagctgataaactctggattgtgctggtt	1372
QY	1373	ccaactctgattgtctgnaaaaaacatcatcatcactgagatacaccatccatccctgtgcgaa	1432
Db	1373	ccaactctgattgtctgnaaaaaacatcatcatcactgagatacaccatccatccctgtgcgaa	1432
QY	1433	ccaagaaatggcnaaataaatactctgggggtcccgatgtgcatactgttgaaagccttgggatttga	1492
Db	1433	ccaagaaatggcnaaataaatactctgggggtcccgatgtgcatactgttgaaagccttgggatttga	1492
QY	1493	ggcagctagatccctacatgacaaagaatgagcccttgtgttactccaagctctctgtgaa	1552
Db	1493	ggcagctagatccctacatgacaaagaatgagcccttgtgttactccaagctctctgtgaa	1552
QY	1553	accacagctctccgtctccacagctgnaagacttgnaatgagacatcaaggaagagctgtgtc	1612
Db	1553	accacagctctccgtctccacagctgnaagacttgnaatgagacatcaaggaagagctgtgtc	1612

```
QY 1613 ccagctggagatgtggtgagctctatagaccatccctctctgcaatcaataaacct 1672
CC |||||||
CC 1613 ccagctggagatgtggtgagctctatagaccatccctctctgcaatcaataaacct 1672
CC |||||||
QY 1673 tgcctgtg 1680
CC |||||||
Db 1673 tgcctgtg 1680

RESULT 3
AAZ87210
ID AAZ87210 standard; cDNA, 1707 BP.
AC AAZ87210;
XX
XX 08-MAY-2000 (first entry)
DE Human NTAP cDNA clone 2799056.
XX
KW Neurotransmission-associated protein; NTAP; odorant-binding protein;
KW neurological disease; Alzheimer's disease; Parkinson's disease;
KW Huntington's disease; cerebral neoplasm; multiple sclerosis;
KW drug screening; gene therapy; antagonist; cancer; AIDS; asthma;
KW Crohn's disease; osteoporosis; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 124..1578
FT /*tag= a
FT /product= "Human neurotransmission-associated protein
PM WO200001821-A2.
XX
XX 13-JAN-2000.
XX
XX 02-JUL-1999; 99MO-US15121.
XX
XX 02-JUL-1998; 98US-0091677.
XX
PA (INCY) INCYTE PHARM INC.
PI Lal P, Tang YF, Yue H, Corley NC, Guegler KJ, Gorgone GA;
PI Baughn MR, Patterson C;
XX
DR WPI: 2000-160770/14.
XX
XX P-PSDB; AAY77126.
XX
XX New human neurotransmission associated proteins, useful for treatment,
XX prevention and diagnosis of neurological disease, e.g. Alzheimer's
XX disease, and antagonists for treating cancer or immune disorders
XX
XX Claim 7; Page 67; 67pp; English.
XX
XX Sequences AAZ87205-287210 represent cDNAs encoding six human
XX neurotransmission-associated proteins (NTAPs, AAY77121-V77126). The
XX present sequence, clone 2799056, encodes a putative odorant-binding
XX protein, and was produced by extension of cDNA fragments isolated from a
XX human nasal polyp tissue cDNA library. The NTAPs are used for treatment
XX or prevention of neurological diseases (e.g., Alzheimer's, Parkinson's
XX or Huntington's diseases), cerebral neoplasms, or multiple sclerosis).
XX They can also be used to raise specific antibodies and to screen for
XX NTAP-encoding nucleic acids (potential (agonists and antagonists).
XX NTAPs, and as a source of therapeutic antagonists (antisenese,
XX triplex-forming or ribozyme molecules). The nucleic acids may also be
XX used as a source of probes and primers for diagnosis or monitoring of
XX NTAP expression in hybridisation/amplification tests, for chromosome
XX mapping and for identifying related sequences, and for gene therapy. NTAP
XX antagonists are used to treat and prevent a wide range of cancers and
XX immune disorders (e.g. AIDS, asthma, Crohn's disease, osteoporosis, viral
```

```
CC or other infections). NTAP antibodies are used to detect NTAPs, for
CC diagnosis or monitoring, as therapeutic antagonists, in competitive drug
CC screens, and for affinity purification of NTAPs from natural sources.
XX
XX Sequence 1707 BP; 406 A; 503 C; 429 G; 369 T; 0 other;
```

```
Query Match 86.8%; Score 1458; DB 21; Length 1707;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1678; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
```

```
QY 1 ggtgtcagagataaagttgtgacttccagaccactgccggagagagagagcggg 60
Db 1 ggtgtcagagataaagttgtgacttccagaccactgccggagagagagagcggg 60
QY 61 ccgagagatccagctgccccaggtctgcaatctgcaatctgcaatctgcaatctg 120
Db 61 ccgagagatccagctgccccaggtctgcaatctgcaatctgcaatctgcaatctg 120
QY 121 aagatgccccgccccgtggacccttccctctctgtgtgtgtgtgtgtgtgtgtgt 180
Db 121 aagatgccccgccccgtggacccttccctctctgtgtgtgtgtgtgtgtgtgtgt 180
QY 181 caagccaccctcagctccactgagttctcaatctcgcgcgaagaagatcaagaaga 240
Db 181 caagccaccctcagctccactgagttctcaatctcgcgcgaagaagatcaagaaga 240
QY 241 ctgacacagagagctggaagacacacacacacacacacacacacacacacacac 300
Db 241 ctgacacagagagctggaagacacacacacacacacacacacacacacacacac 300
QY 301 agtgcacatcgaggaaaagccagccgagag-atccctgtgtgtgtgtgtgtgtgtgt 359
Db 301 agtgcacatcgaggaaaagccagccgagag-atccctgtgtgtgtgtgtgtgtgtgt 359
QY 360 gtccgtgaagacatcatctgtgtgaagttcatcacagatcaatcatccacatgagtg 419
Db 360 gtccgtgaagacatcatctgtgtgaagttcatcacagatcaatcatccacatgagtg 419
QY 420 aagccctgggcaatgacacagagagctgtatgtaagaatccccctggagatgtgtga 479
Db 420 aagccctgggcaatgacacagagagctgtatgtaagaatccccctggagatgtgtga 479
QY 480 ttcaacacgccccctgtgtcaagacacatctgtgagttccacatgagacatgagcccaagcc 539
Db 480 ttcaacacgccccctgtgtcaagacacatctgtgagttccacatgagacatgagcccaagcc 539
QY 481 ttcaacacgccccctgtgtcaagacacatctgtgagttccacatgagacatgagcccaagcc 540
Db 481 ttcaacacgccccctgtgtcaagacacatctgtgagttccacatgagacatgagcccaagcc 540
QY 540 accatccgcatggaacacacagatgcaagtgccccacacccgctgtctcctcagtgactgtgc 599
Db 540 accatccgcatggaacacacagatgcaagtgccccacacccgctgtctcctcagtgactgtgc 599
QY 600 accagccatggagagctgtgcatccactgtgtgtaagaatctctctcctcgtgtgagcc 659
Db 600 accagccatggagagctgtgcatccactgtgtgtaagaatctctctcctcgtgtgagcc 659
QY 660 tttagtaagcaggtcatgtaacatctctatgtgcatccctgtgccaatctagtgaaataacacag 719
Db 660 tttagtaagcaggtcatgtaacatctctatgtgcatccctgtgccaatctagtgaaataacacag 719
QY 720 ctgtgtccctgtgagaggtctctcctcaatgtgcatgtatgacagactctgtgagctgtgtg 779
Db 720 ctgtgtccctgtgagaggtctctcctcaatgtgcatgtatgacagactctgtgagctgtgtg 779
QY 780 aagtgcccatcttccctcagacatgacggtctgtgagttgacttctgtatctctgcacatc 839
Db 780 aagtgcccatcttccctcagacatgacggtctgtgagttgacttctgtatctctgcacatc 839
QY 840 aagtggtacacatcatgactctacactgtggtggccaagtgtgtgactcaagaggaagtg 899
Db 840 aagtggtacacatcatgactctacactgtggtggccaagtgtgtgactcaagaggaagtg 899
QY 900 accaagtggtcaataactctgagcttccctgtacaatgtccacccctgtgacacatcccg 959
Db 900 accaagtggtcaataactctgagcttccctgtacaatgtccacccctgtgacacatcccg 959
```



Db 901 accaagtggtcaataactctgcagcttccctgcacatgcccacgtgacaacatccg 960  
QY 960 tttagagctatcgttgtagtagacgctggtgaaagctgcagctgctgctctctcca 1019  
Db 961 tttagagctatcgttgtagtagacgctggtgaaagctgcagctgctgctctctcca 1020  
QY 1020 gaagaatcatgtctctgtgtagctctgtctctcctgcgaagtgccatcgctggaagta 1079  
Db 1021 gaagaatcatgtctctgtgtagctctgtctctcctgcgaagtgccatcgctggaagta 1080  
QY 1080 agcatcgggctgtagcaatgaaaggtctgagataagcttgagatactaccagatctgaaag 1139  
Db 1081 agcatcgggctgtagcaatgaaaggtctgagataagcttgagatactaccagatctgaaag 1140  
QY 1140 atccctaactcaagacacctcccgagtttttatagaccaggccatgccaagtgtgccaa 1199  
Db 1141 atccctaactcaagacacctcccgagtttttatagaccaggccatgccaagtgtgccaa 1200  
QY 1200 ctgacgtgctggaagtgcttccctcccaagtgaaagccctcgcccttgttcaacctgggc 1259  
Db 1201 ctgacgtgctggaagtgcttccctcccaagtgaaagccctcgcccttgttcaacctgggc 1260  
QY 1260 atcgaaagcagctcggaaagctcaagttttaacccaaggtgaccaacttatactcaactg 1319  
Db 1261 atcgaaagcagctcggaaagctcaagttttaacccaaggtgaccaacttatactcaactg 1320  
QY 1320 aataaatacagctctgtagtggtcccgctgtagtaactctggagttgctggttccaact 1379  
Db 1321 aataaatacagctctgtagtggtcccgctgtagtaactctggagttgctggttccaact 1380  
QY 1380 gatgtctgtaaaaaacatcatcactgtagatcatcactcatcctgtctgcggaacgaat 1439  
Db 1381 gatgtctgtaaaaaacatcatcactgtagatcatcactcatcctgtctgcggaacgaat 1440  
QY 1440 ggcaaatagaatctgggggtcccaagtgtcatctggtgaaggtctggagttcgagagcagt 1499  
Db 1441 ggcaaatagaatctgggggtcccaagtgtcatctggtgaaggtctggagagcagt 1500  
QY 1500 gagtcctcaactgacaaagagatgccttgtgttactccagccctctgttggaacccaagc 1559  
Db 1501 gagtcctcaactgacaaagagatgccttgtgttactccagccctctgttggaacccaagc 1560  
QY 1560 tctcctgtctcccaagtgaagacttgtagtgcaagccatcagggaaagctggtgccagctg 1619  
Db 1561 tctcctgtctcccaagtgaagacttgtagtgcaagccatcagggaaagctggtgccagctg 1620  
QY 1620 ggaagtaggggtgtagagcttatagaccatccctctctgcaatcaataaacaactgtcctgt 1679  
Db 1621 ggaagtaggggtgtagagcttatagaccatccctctctgcaatcaataaacaactgtcctgt 1680  
QY 1680 g 1680  
Db 1681 g 1681

## RESULT 4

AAA37057  
ID AAA37057 standard; cDNA; 1636 BP.

AC AAA37057;

DT 08-AUG-2000 (first entry)

DE Human PRO1357 (UNG706) cDNA sequence SEQ ID NO:127.

KW Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;  
transmembrane; secretion; immunoadhesion; pharmaceutical; screening;  
ss.

OS Homo sapiens.

PN WO200012708-A2.

XX

PD 09-MAR-2000.  
XX  
PF 01-SEP-1999; 99WO-US20111.  
XX  
PR 01-SEP-1998; 98US-0098716.  
PR 01-SEP-1998; 98US-0098749.  
PR 01-SEP-1998; 98US-0098750.  
PR 02-SEP-1998; 98US-0098803.  
PR 02-SEP-1998; 98US-0098821.  
PR 02-SEP-1998; 98US-0098843.  
PR 09-SEP-1998; 98US-0099536.  
PR 09-SEP-1998; 98US-0099596.  
PR 09-SEP-1998; 98US-0099598.  
PR 09-SEP-1998; 98US-0099602.  
PR 09-SEP-1998; 98US-0099642.  
PR 10-SEP-1998; 98US-0099741.  
PR 10-SEP-1998; 98US-0099754.  
PR 10-SEP-1998; 98US-0099763.  
PR 10-SEP-1998; 98US-0099792.  
PR 10-SEP-1998; 98US-0099808.  
PR 10-SEP-1998; 98US-0099812.  
PR 10-SEP-1998; 98US-0099815.  
PR 10-SEP-1998; 98US-0099816.  
PR 15-SEP-1998; 98US-0100385.  
PR 15-SEP-1998; 98US-0100388.  
PR 15-SEP-1998; 98US-0100390.  
PR 16-SEP-1998; 98US-0100584.  
PR 16-SEP-1998; 98US-0100627.  
PR 16-SEP-1998; 98US-0100661.  
PR 16-SEP-1998; 98US-0100662.  
PR 16-SEP-1998; 98US-0100664.  
PR 17-SEP-1998; 98US-0100683.  
PR 17-SEP-1998; 98US-0100684.  
PR 17-SEP-1998; 98US-0100710.  
PR 17-SEP-1998; 98US-0100711.  
PR 17-SEP-1998; 98US-0100919.  
PR 17-SEP-1998; 98US-0100930.  
PR 18-SEP-1998; 98US-0100848.  
PR 18-SEP-1998; 98US-0100849.  
PR 18-SEP-1998; 98US-0101014.  
PR 18-SEP-1998; 98US-0101068.  
PR 18-SEP-1998; 98US-0101071.  
PR 22-SEP-1998; 98US-0101279.  
PR 23-SEP-1998; 98US-0101471.  
PR 23-SEP-1998; 98US-0101472.  
PR 23-SEP-1998; 98US-0101474.  
PR 23-SEP-1998; 98US-0101475.  
PR 23-SEP-1998; 98US-0101476.  
PR 23-SEP-1998; 98US-0101477.  
PR 23-SEP-1998; 98US-0101479.  
PR 24-SEP-1998; 98US-0101738.  
PR 24-SEP-1998; 98US-0101741.  
PR 24-SEP-1998; 98US-0101743.  
PR 24-SEP-1998; 98US-0101915.  
PR 24-SEP-1998; 98US-0101916.  
PR 29-SEP-1998; 98US-0102207.  
PR 29-SEP-1998; 98US-0102240.  
PR 29-SEP-1998; 98US-0102307.  
PR 29-SEP-1998; 98US-0102309.  
PR 29-SEP-1998; 98US-0102331.  
PR 30-SEP-1998; 98US-0102484.  
PR 30-SEP-1998; 98US-0102487.  
PR 30-SEP-1998; 98US-0102570.  
PR 30-SEP-1998; 98US-0102571.  
PR 01-OCT-1998; 98US-0102684.  
PR 01-OCT-1998; 98US-0102687.  
PR 02-OCT-1998; 98US-0102965.  
PR 06-OCT-1998; 98US-0103258.  
PR 06-OCT-1998; 98US-0103314.  
PR 07-OCT-1998; 98US-0103315.  
PR 07-OCT-1998; 98US-0103328.  
PR 07-OCT-1998; 98US-0103395.

[illegible]

```

Db 903 acatccgcttcagcctcatcgtgagtcagagcgtggtgaaagctgacgtgctgtgtgc 962
Qy 1012 tcttcacgaagaattcatggtccctgttgactctgtcttcttaagagtgccactgcgc 1071
Db 963 tcttcacgaagaattcatggtccctgttgactctgtcttcttaagagtgccactgcgc 1022
Qy 1072 tgaagtcagacgtcggcgtgatacaatgaaagcgtcgaataagctggatctaccaga 1131
Db 1023 tgaagtcagacgtcggcgtgatacaatgaaagcgtcgaataagctggatctaccaga 1082
Qy 1132 tctgtaagatccttaactcaagacactcccgagtttcttataagacaaagccatgcgaag 1191
Db 1083 tctgtaagatccttaactcaagacactcccgagtttcttataagacaaagccatgcgaag 1142
Qy 1192 tggcccaactgactgtctgtgaaagtgttccctcgaatgaaacccctccgcttctgttca 1251
Db 1143 tggcccaactgactgtctgtgaaagtgttccctcgaatgaaacccctccgcttctgttca 1202
Qy 1252 cccttgagatcgaaagcagcctcgaaagctcaagtttatacaccaaagtgacaactataac 1311
Db 1203 cccttgagatcgaaagcagcctcgaaagctcaagtttatacaccaaagtgacaactataac 1262
Qy 1312 tcaacttgaataacatacagcctctgatacggatccagctgataactctggatgtgctggt 1371
Db 1263 tcaacttgaataacatacagcctctgatacggatccagctgataactctggatgtgctggt 1322
Qy 1372 tccaacctgagtgcttctgaaaaacatcatcactgacagatatacctaccctcctgcgcga 1431
Db 1323 tccaacctgagtgcttctgaaaaacatcatcactgacagatatacctaccctcctgcgcga 1382
Qy 1432 accaagaatgacaaatlaagatcttggtgtcccaagctgatacttgatgaagccttgagatcg 1491
Db 1383 accaagaatgacaaatlaagatcttggtgtcccaagctgatacttgatgaagccttgagatcg 1442
Qy 1492 aggcagcgtgagtcctcactgacacaaagatgcccctgtgtcttaactccagcctcctgtgga 1551
Db 1443 aggcagcgtgagtcctcactgacacaaagatgcccctgtgtcttaactccagcctcctgtgga 1502
Qy 1552 aaccagcgtcctctgtctcccaagtaagacttgatctgagcagcactcaagggaagcgtggt 1611
Db 1503 aaccagcgtcctctgtctcccaagtaagacttgatctgagcagcactcaagggaagcgtggt 1562
Qy 1612 cccaagctggagatggtgtgtgagctctatagacatccctctctgcaatacaataaacac 1671
Db 1563 cccaagctggagatggtgtgtgagctctatagacatccctctctgcaatacaataaacac 1622
Qy 1672 ttgctctgtg 1680
Db 1623 ttgctctgtg 1631

```

```

RESULT 5
AAS46064
ID AAS46064 standard; cDNA; 1636 BP.
AC AAS46064;

```

```

XX 18-DEC-2001 (first entry)

```

```

DE Human DNA encoding PRO polypeptide sequence #140.

```

```

XX PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep; ss;
KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder;
KW PCR primer.

```

```

XX Homo sapiens.

```

```

XX WO200168848-A2.

```

```

XX 20-SEP-2001.

```

```

XX

```

```

PF 28-FEB-2001; 2001WO-US06520.
XX
PR 01-MAR-2000; 2000WO-US05601.
PR 02-MAR-2000; 2000WO-US05841.
PR 03-MAR-2000; 2000US-187202P.
PR 06-MAR-2000; 2000US-186968P.
PR 14-MAR-2000; 2000US-189320P.
PR 14-MAR-2000; 2000US-189320P.
PR 15-MAR-2000; 2000WO-US06884.
PR 21-MAR-2000; 2000US-190828P.
PR 21-MAR-2000; 2000US-191007P.
PR 21-MAR-2000; 2000US-191048P.
PR 21-MAR-2000; 2000US-191314P.
PR 28-MAR-2000; 2000US-192655P.
PR 29-MAR-2000; 2000US-193032P.
PR 29-MAR-2000; 2000US-193053P.
PR 30-MAR-2000; 2000WO-US08439.
PR 04-APR-2000; 2000US-194449P.
PR 04-APR-2000; 2000US-194647P.
PR 11-APR-2000; 2000US-195975P.
PR 11-APR-2000; 2000US-196000P.
PR 11-APR-2000; 2000US-196187P.
PR 11-APR-2000; 2000US-196690P.
PR 11-APR-2000; 2000US-196820P.
PR 18-APR-2000; 2000US-198121P.
PR 18-APR-2000; 2000US-198585P.
PR 25-APR-2000; 2000US-199397P.
PR 25-APR-2000; 2000US-199550P.
PR 25-APR-2000; 2000US-199654P.
PR 03-MAY-2000; 2000US-201516P.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 05-JUN-2000; 2000US-209832P.
PR 28-JUL-2000; 2000WO-US20710.
PR 22-AUG-2000; 2000US-064484P.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000WO-US34956.
XX
XX (GENTH ) GENENTECH INC.
XX
XX Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
XX
XX WPI; 2001-602746/68.
XX
XX P-PSDB; AAU29163.
XX
XX Novel nucleic acids encoding PRO polypeptides, used to diagnose the
PT presence of tumours, such as prostate and breast tumours, in mammals and
PT to screen for modulators of the compounds -
XX
XX Claim 2: Fig 279; 774pp; English.
XX
XX Sequences AAS45925-AAS46231 represent DNA molecules encoding and PCR
CC primers for PRO polypeptides of the invention. The sequences of the
CC invention can be used to detect the presence of a tumour in a mammal by
CC comparing the level of expression of a PRO polypeptide in a test sample
CC of cells from the animal and a control sample of normal cells, whereby a
CC higher level of expression in the test sample indicates the presence of a
CC tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep,
CC pigs, goats and rabbits but are preferably human. The polypeptides can be
CC used to stimulate tumour necrosis factor (TNF) alpha release from human
CC blood, when contacted with it. A specific polypeptide can be used to
CC stimulate the proliferation or differentiation of chondrocyte cells. The
CC PRO proteins can be used to determine the presence of tumours and also
CC susceptibility to tumour development, particularly adrenal, lung, colon,
CC breast, prostate, rectal, cervical, or liver tumours, in mammalian
CC subjects. The oligonucleotide probes specific for the PRO nucleic acids
CC can be used for genetic analysis of individuals with genetic disorders.
XX

```

Sequence 1636 BP; 374 A; 492 C; 411 G; 359 T; 0 other:

Query Match 86.7%; Score 1457; DB 22; Length 1636;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1627; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```
QY 53 ggaagcgagccagagatccagcgtgcccaggtctgcatccctgcaactgtgctcccttga 112
Db 3 ggaagcgagccagagatccagcgtgcccaggtctgcatccctgcaactgtgctcccttga 62
QY 113 caacttggaagatgagccggtccgttggaaccttcaacctctgtgtgtctgagcagca 172
Db 63 caacttggaagatgagccggtccgttggaaccttcaacctctgtgtgtgtgagcagca 122
QY 173 ccttgatccaaagccaaacctcaagctcccaatgcatctcaatctctgagcccaaaatgcatca 232
Db 123 ccttgatccaaagccaaacctcaagctcccaatgcatctcaatctctgagcccaaaatgcatca 182
QY 233 aagaaaagctgacacagagagctgaaagacacaaacgacacagacatccctgacagcgtcgc 292
Db 183 aagaaaagctgacacagagagctgaaagacacaaacgacacagacatccctgacagcgtcgc 242
QY 293 cgcctcctcaagctgacagcgtggaagacagacagcagcagcagcagcagcagcagcagcagc 351
Db 243 cgcctcctcaagctgacagcgtggaagacagacagcagcagcagcagcagcagcagcagcagc 302
QY 352 tgaacacacgctctgacacacatcatctgctgaaagctgacacacagacacacacacacacac 411
Db 303 tgaacacacgctctgacacacatcatctgctgaaagctgacacacagacacacacacacacac 362
QY 412 tgcaggtgaagccctccgacacaaatgacagagcgtgcatgtaagaatccctctgacatg 471
Db 363 tgcaggtgaagccctccgacacaaatgacagagcgtgcatgtaagaatccctctgacatg 422
QY 472 tgcgtgtaattcaaacagcccttggccaagacatcgtggaattccacatgacagctagg 531
Db 423 tgcgtgtaattcaaacagcccttggccaagacatcgtggaattccacatgacagctagg 482
QY 532 cccaagacacacatccgcatgtagacacacagctgcaagtggcccccacgctggtccctcagtg 591
Db 483 cccaagacacacatccgcatgtagacacacagctgcaagtggcccccacgctggtccctcagtg 542
QY 592 actgtgccaacagacacagagcctgctgcatccaactgctgcatgaagcttcccttccctgg 651
Db 543 actgtgccaacagacacagagcctgctgcatccaactgctgcatgaagcttcccttccctgg 602
QY 652 tgaacgccttaagctaaagcaggtcaatgaacctcctagtgccatccctgcccacatcagtaga 711
Db 603 tgaacgccttaagctaaagcaggtcaatgaacctcctagtgccatccctgcccacatcagtaga 662
QY 712 aaaaacgacgtgtcccgtagatcgaggtccttcaatgagcatgtagacacacacacacacac 771
Db 663 aaaaacgacgtgtcccgtagatcgaggtccttcaatgagcatgtagacacacacacacacac 722
QY 772 agctgtgtgaaggtgcccatttccctcagcaatgacacgctgaggttgaacctgtatc 831
Db 723 agctgtgtgaaggtgcccatttccctcagcaatgacacgctgaggttgaacctgtatc 782
QY 832 ctgcatcaaggtgtagacacatcagcttcaacacgctgaggttgaacctgtatc 891
Db 783 ctgcatcaaggtgtagacacatcagcttcaacacgctgaggttgaacctgtatc 842
QY 892 gaaaggtgaccaaagtgttcaataactgtagcagcttccctgacaaatgcccaccccggaaca 951
Db 843 gaaaggtgaccaaagtgttcaataactgtagcagcttccctgacaaatgcccaccccggaaca 902
QY 952 acatccgcttcaagcctcagcgtgagtcagagacgtagtgaagctgcagctgctgtgc 1011
Db 903 acatccgcttcaagcctcagcgtgagtcagagacgtagtgaagctgcagctgctgtgc 962
QY 1012 tctctccagaagaatcatggtccctgtgagctctgtcttctcgaaggtgccatcggc 1071
Db 1012 tctctccagaagaatcatggtccctgtgagctctgtcttctcgaaggtgccatcggc 1071
```

```
Db 963 tctctccagaagaatcatggtccctgtgagctctgtcttctcgaaggtgccatcggc 1022
QY 1072 tgaagtcagaagatcggtgctgataatgaataaagctgtagaataagctggaattcaccaga 1131
Db 1023 tgaagtcagaagatcggtgctgataatgaataaagctgtagaataagctggaattcaccaga 1082
QY 1132 tctgtgaagatcctaactcagagacacatcccgagtttttataagaccaaagcagctgcccag 1191
Db 1083 tctgtgaagatcctaactcagagacacatcccgagtttttataagaccaaagcagctgcccag 1142
QY 1192 tggcccaactgatactgctgtggaagtgttccctccaggtgaagccctccgcttggta 1251
Db 1143 tggcccaactgatactgctgtggaagtgttccctccaggtgaagccctccgcttggta 1202
QY 1252 cccctggacatcgaagacagcagctcggaagctgatttaccaccaaaggtgacaaattatc 1311
Db 1203 cccctggacatcgaagacagcagctcggaagctgatttaccaccaaaggtgacaaattatc 1262
QY 1312 tcaacttgaataacatcagctgatacagatccagctgatactggaactggaattgctggt 1371
Db 1263 tcaacttgaataacatcagctgatacagatccagctgatactggaactggaattgctggt 1322
QY 1372 tccaactgatagttctggaataacatcagctgatactggaactggaattgctggt 1431
Db 1323 tccaactgatagttctggaataacatcagctgatactggaactggaattgctggt 1382
QY 1432 accagaatggaataatgaatcgtgggtccagtgatctggttgaagccttgagatc 1491
Db 1383 accagaatggaataatgaatcgtgggtccagtgatctggttgaagccttgagatc 1442
QY 1492 agcagactgagctcctcagcagacagagagcctgtgtcttcaactcagcctcctgtgga 1551
Db 1443 agcagactgagctcctcagcagacagagagcctgtgtcttcaactcagcctcctgtgga 1502
QY 1552 aacccaagctcctcgtctccagtgaaagactggaatggcagccatcagaagagcgtggt 1611
Db 1503 aacccaagctcctcgtctccagtgaaagactggaatggcagccatcagaagagcgtggt 1562
QY 1612 cccaagctggaatgaggtgtagagctctatagacacatccctcctgtgcaataaataaac 1671
Db 1563 cccaagctggaatgaggtgtagagctctatagacacacatccctcctgtgcaataaataaac 1622
QY 1672 ttgctgtg 1680
Db 1623 ttgctgtg 1631
```

RESULT 6  
AAF92096  
ID AAF92096 standard; cDNA; 1636 BP.  
XX  
AC AAF92096;  
XX  
DT 15-MAY-2001 (first entry)  
XX  
DE Human PRO1357 cDNA.  
XX  
KW Human; PRO protein; mapping; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200116318-A2.  
XX  
PD 08-MAR-2001.  
XX  
PE 24-AUG-2000; 2000MO-US23328.  
XX  
PR 01-SEP-1999; 99MO-US20111.  
PR 15-SEP-1999; 99MO-US21090.  
PR 07-DEC-1999; 99US-0169495.  
PR 09-DEC-1999; 99US-0170262.  
PR 11-JAN-2000; 2000US-0175481.  
PR 18-FEB-2000; 2000MO-US04341.

PR 18-FEB-2000; 2000MO-US04342.  
 PR 22-FEB-2000; 2000MO-US04414.  
 PR 01-MAR-2000; 2000MO-US05601.  
 PR 03-MAR-2000; 2000MO-0187202.  
 PR 25-APR-2000; 2000MO-0193997.  
 PR 22-MAY-2000; 2000MO-US14042.  
 PR 05-JUN-2000; 2000MO-0209832.  
 XX  
 PA (GENTH ) GENENTECH INC.  
 XX  
 PI Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;  
 PI Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;  
 XX  
 DR WPI; 2001-183260/18.  
 DR P-PSDB; AAB87564.  
 XX  
 PT Eighty four nucleic acids encoding PRO polypeptides, useful in  
 PT molecular biology, including use as hybridization probes, and in  
 PT chromosome and gene mapping.  
 XX  
 PS Claim 2; Fig 77; 278pp; English.  
 CC The present sequence is the coding sequence for a human PRO polypeptide  
 CC (secreted and transmembrane). The PRO protein, and PRO agonists, PRO  
 CC antagonists or anti-PRO antibodies are useful for preparation of a  
 CC medicament useful in the treatment of a condition which is responsive to  
 CC the PRO protein, agonists, antagonists or anti-PRO antibodies. The PRO  
 CC protein may also be employed as molecular weight markers for protein  
 CC electrophoresis. The PRO coding sequence has applications in molecular  
 CC biology, including use as hybridisation probes, and in chromosome and  
 CC gene mapping.  
 XX  
 SO Sequence 1636 BP; 374 A; 492 C; 411 G; 359 T; 0 other;

Query Match 86.7%; Score 1457; DB 22; Length 1636;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1627; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 53 ggaagcgccgagagatccagcgtgccaggtctgcatccctgacgtctgctcctccta 112  
 DB 3 ggaagcgccgagagatccagcgtgccaggtctgcatccctgacgtctgctcctccta 62  
 QY 113 caactgggaagatgagcgccgtgagacctcaacctctctgtgtgttgctgagcgca 172  
 DB 63 caactgggaagatgagcgccgtgagacctcaacctctctgtgtgttgctgagcgca 122  
 QY 173 ccttgatccaagccacctcagctcccaatgcatctcattcctcgcgcgaagatcata 232  
 DB 123 ccttgatccaagccacctcagctcccaatgcatctcattcctcgcgcgaagatcata 182  
 QY 233 aagaaagctgacacagagagctgaaagacacaaagccacacagatcctgcaagctgc 292  
 DB 183 aagaaagctgacacagagagctgaaagacacaaagccacacagatcctgcaagctgc 242  
 QY 293 cactgtcagatcccatgagcggaagacagccggaagc-attcctgtgctgggagactgg 351  
 DB 243 cactgtcagatcccatgagcggaagacagccggaagc-attcctgtgctgggagactgg 302  
 QY 352 tgaacacgctcctgaaagcatatctgtgctgaaagctcaacagtaacatcctcagc 411  
 DB 303 tgaacacgctcctgaaagcatatctgtgctgaaagctcaacagtaacatcctcagc 362  
 QY 412 tgaagtgaaagcctcgcgaatgacacagagctgctagtcaagaatccctcctgacatg 471  
 DB 363 tgaagtgaaagcctcgcgaatgacacagagctgctagtcaagaatccctcctgacatg 422  
 QY 472 tggctgattcaacacgcccctgtgtaagaacatctgtgagttccatgacacagactgaag 531  
 DB 423 tggctgattcaacacgcccctgtgtaagaacatctgtgagttccatgacacagactgaag 482  
 QY 532 cccaagccaacatccgcatggaacacagtgcaagtgcccccgcgctgtcctcagtg 591  
 DB 1632 cccaagccaacatccgcatggaacacagtgcaagtgcccccgcgctgtcctcagtg

DB 483 cccaagccaacatccgcatggaacacagtgcaagtgcccccgcgctgtcctcagtg 542  
 QY 592 actgtgtccacccagccatgtggagcctcgcgcatccaactgtgtcataagctctccttcgg 651  
 DB 543 actgtgtccacccagccatgtggagcctcgcgcatccaactgtgtcataagctctccttcgg 602  
 QY 652 tgaacgcttaagcagcagctatgaacatcctcctagtgcacatccctgcgaatctagtga 711  
 DB 603 tgaacgcttaagcagcagctatgaacatcctcctagtgcacatccctgcgaatctagtga 662  
 QY 712 aaaaacagctgtgtcccgatgacaggtcctcctcaatgagcagatgtaagacatccttcgc 771  
 DB 663 aaaaacagctgtgtcccgatgacaggtcctcctcaatgagcagatgtaagacatccttcgc 722  
 QY 772 agctgtgtaagtggtcccatcttcctcaagatgagccgctgtgagtttgacctgtatc 831  
 DB 723 agctgtgtaagtggtcccatcttcctcaagatgagccgctgtgagtttgacctgtatc 782  
 QY 832 ctgccaataaggtgacacacatcagctcagctgagggccaaagtgttgagactcaag 891  
 DB 783 ctgccaataaggtgacacacatcagctcagctgagggccaaagtgttgagactcaag 842  
 QY 892 gaaagtgaaagtggttcaataactctgcagctcctcctgacaatgcccacccctgaca 951  
 DB 843 gaaagtgaaagtggttcaataactctgcagctcctcctgacaatgcccacccctgaca 902  
 QY 952 acatcccgcttcagccatcctgagatcagagcgtgtgtaagagtgagtggtcgtcgtgc 1011  
 DB 903 acatcccgcttcagccatcctgagatcagagcgtgtgtaagagtgagtggtcgtcgtgc 962  
 QY 1012 tctctcagaagaatcatggtctctgtgacatctgtctcctgagatgcccacgcgc 1071  
 DB 963 tctctcagaagaatcatggtctctgtgacatctgtctcctgagatgcccacgcgc 1022  
 QY 1072 tgaagtcagagcatcgggctgataatgaaagctgcagataagcgtggagctaccacaga 1131  
 DB 1023 tgaagtcagagcatcgggctgataatgaaagctgcagataagcgtggagctaccacaga 1082  
 QY 1132 tctgtaagatccctaactcagagacatcccgagtttttataagccaagccatgccaag 1191  
 DB 1083 tctgtaagatccctaactcagagacatcccgagtttttataagccaagccatgccaag 1142  
 QY 1192 tggcccaactgacgtgctgcggaagtgttcccccagtgaaagccctccgctctgttca 1251  
 DB 1143 tggcccaactgacgtgctgcggaagtgttcccccagtgaaagccctccgctctgttca 1202  
 QY 1252 cctctggcatcgaagcagctcggaagctcaagttttacacccaaggttgacccaactatac 1311  
 DB 1203 cctctggcatcgaagcagctcggaagctcaagttttacacccaaggttgacccaactatac 1262  
 QY 1312 tcaacttgaataacatcagctctgatacagatcagatgaaactctgtgattgtgctgtg 1371  
 DB 1263 tcaacttgaataacatcagctctgatacagatcagatgaaactctgtgattgtgctgtg 1322  
 QY 1372 tccaacttgatgttctgaaanaaacatcatcagctagatcattcaactcattcctgtcgca 1431  
 DB 1323 tccaacttgatgttctgaaanaaacatcatcagctagatcattcaactcattcctgtcgca 1382  
 QY 1432 accaagaatggaataatgaatctggtgtccagtgatcatctgtgtaagagccttggatcgc 1491  
 DB 1383 accaagaatggaataatgaatctggtgtccagtgatcatctgtgtaagagccttggatcgc 1442  
 QY 1492 aggcagctgagttcctcagcagcaagatgcccctgtgttactcagcctcctctgtgga 1551  
 DB 1443 aggcagctgagttcctcagcagcaagatgcccctgtgttactcagcctcctctgtgga 1502  
 QY 1552 aaccagctctcctgtctccagtgaaagcttgatgagcagcacaaggaagctggggt 1611  
 DB 1503 aaccagctctcctgtctccagtgaaagcttgatgagcagcacaaggaagctggggt 1562  
 QY 1612 cccaagctggaatgataaggtgtgagctctatagacatccctctctgcaataataaacac 1671  
 DB 1563 cccaagctggaatgataaggtgtgagctctatagacatccctctctgcaataataaacac 1622

QY 1672 ttgcctgtg 1680  
|||||||  
Db 1623 ttgcctgtg 1631

## RESULT 7

AAFS4293  
ID AAF54293 standard; DNA; 1636 BP.

AC AAF54293;

DT 02-APR-2001 (first entry)

DE DNA encoding protein of the invention #36.

KW Secreted; transmembrane; gene therapy; ss.

OS Unidentified.

XX WO200078961-A1.

XX PD 28-DEC-2000.

XX PF 18-FEB-2000; 2000WO-US04342.

XX PR 23-JUN-1999; 99US-0141037.

XX PR 20-JUL-1999; 99US-014758.

XX PR 26-JUL-1999; 99US-0145698.

XX PR 01-SEP-1999; 99WO-US20111.

XX PR 29-OCT-1999; 99US-0162506.

XX PR 30-NOV-1999; 99WO-US28313.

XX PR 02-DEC-1999; 99WO-US28551.

XX PR 16-DEC-1999; 99WO-US30095.

XX PR 05-JAN-2000; 2000WO-US00219.

XX PR 06-JAN-2000; 2000WO-US00376.

XX PA (GENTH) GENENTECH INC.

XX PI Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;

XX PI Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;

XX PI Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;

XX PI Watanabe CK, Williams PM, Wood WI;

XX DR WPI; 2001-071395/08.

XX PT Secreted and transmembrane proteins and nucleic acids designated PRO,

XX PT useful as hybridization probes, in chromosome and gene mapping and gene

XX PT therapy.

XX PS Claim 2; Fig 71; 787pp; English.

XX CC The present invention relates to secreted and transmembrane proteins.

XX CC These proteins and the DNA encoding them may be used as hybridization

XX CC probes, in chromosome and gene mapping and in the generation of

XX CC anti-sense RNA and DNA. They may also be used to generate either

XX CC transgenic animals or knockout animals which are in turn useful for

XX CC development and screening of therapeutically useful reagents.

XX CC The nucleic acids may also be used in gene therapy.

XX SO Sequence 1636 BP; 374 A; 492 C; 411 G; 359 T; 0 other;

Query Match 86.7%; Score 1457; DB 22; Length 1636;  
Best Local Similarity 99.9%; Pred. No. 0;

Matches 1627; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 53 gggagggccagagactcagcgtgccaggtctggcactcgtgactgtgacctcgtga 112

Db 3 gggagggccagagactcagcgtgccaggtctggcactcgtgactgtgacctcgtga 62

QY 113 cacttgggaagatggcggcgctgacttaacccttctgtgtgttctgtgagca 172

|||||||

Db 63 cacttgggaagatggcggcgctgacttaacccttctgtgtgttctgtgagca 122

QY 173 ccttgatccaagcacccttcagctccactgacgttccatctctcctcggccaaagatca 232

Db 123 ccttgatccaagcacccttcagctccactgacgttccatctctcctcggccaaagatca 182

QY 233 aagaagaagctgaacagagctgaagagcacaacagccacagcatctcgaacagctgc 292

Db 183 aagaagaagctgaacagagctgaagagcacaacagccacagcatctcgaacagctgc 242

QY 293 cgtctcagtgatgcacatcgggaaagccagccgagag-atccctgtgtcgggagccttg 351

Db 243 cgtctcagtgatgcacatcgggaaagccagccgagag-atccctgtgtcgggagccttg 302

QY 352 tgaacacgctctgaagcacatcatctgtgtgaaggtcatcagcttaacatctcagc 411

Db 303 tgaacacgctctgaagcacatcatctgtgtgaaggtcatcagcttaacatctcagc 362

QY 412 tgcaggtgaagccctcggcacaatgaacagagctgtcgtatcaagatcccccgtgacatg 471

Db 363 tgcaggtgaagccctcggcacaatgaacagagctgtcgtatcaagatcccccgtgacatg 422

QY 472 tggctggtatccaacagccctgtgtcaagacatcgtgtgagttccaatgaagactgag 531

Db 423 tggctggtatccaacagccctgtgtcaagacatcgtgtgagttccaatgaagactgag 482

QY 532 cccaagcacaacatccgcatgagacacagatgcaaggtggcccaacccgctgtctcagtg 591

Db 483 cccaagcacaacatccgcatgagacacagatgcaaggtggcccaacccgctgtctcagtg 542

QY 592 actgtgccacccaagccttggaagcctgcgcatccaactgtgtatgaagctctcctcctg 651

Db 543 actgtgccacccaagccttggaagcctgcgcatccaactgtgtatgaagctctcctcctg 602

QY 652 tgaacgcttagtgaagcagatgaacacccctcctagtgtccatccctgccaatcagtg 711

Db 603 tgaacgcttagtgaagcagatgaacacccctcctagtgtccatccctgccaatcagtg 662

QY 712 aaacacagctgtgtcccgatcgagcgtctccttaagagcatgatatcagaactcctcgc 771

Db 663 aaacacagctgtgtcccgatcgagcgtctccttaagagcatgatatcagaactcctcgc 722

QY 772 agctgtgaaggtgccatcttccctcaagatgaacgctgtgagttgaactctgtatc 831

Db 723 agctgtgaaggtgccatcttccctcaagatgaacgctgtgagttgaactctgtatc 782

QY 832 ctggcatcaaggtgaacacatctagctctcctcctcctcctcctcctcctcctcctcctc 891

Db 783 ctggcatcaaggtgaacacatctagctctcctcctcctcctcctcctcctcctcctcctc 842

QY 892 gaaagtgaccagtggttcaataactctgacgttcccttgacatgcccacccctgaca 951

Db 843 gaaagtgaccagtggttcaataactctgacgttcccttgacatgcccacccctgaca 902

QY 952 acatccgttcaagctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1011

Db 903 acatccgttcaagctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 962

QY 1012 tctctcagaagaattcaatgttccctgtgtgacgttgccttccctcctcctcctcctcctcctc 1071

Db 963 tctctcagaagaattcaatgttccctgtgtgacgttgccttccctcctcctcctcctcctcctc 1022

QY 1072 tgaagtgacatcgggtgtgatacaatgaagaagctgtcagataagctgtgataccaga 1131

Db 1023 tgaagtgacatcgggtgtgatacaatgaagaagctgtcagataagctgtgataccaga 1082

QY 1132 tctgtgaagatcttaactcaggaacactccgaagttttatagaaccaagccatgacaag 1191

Db 1083 tctgtgaagatcttaactcaggaacactccgaagttttatagaaccaagccatgacaag 1142

QY 1192 tggcccaactgacgtgtcgtgaaggtgttccctcagtgtagccctcgcgccttcttca 1251

Db 1143 tggcccaactgacgtgtcgtgaaggtgttccctcagtgtagccctcgcgccttcttca 1202



RESULT	9
AA572707	
ID	AA572707 standard; CDNA; 2121 BP.
XX	
XX	
AC	AA572707;
XX	
DT	13-FEB-2002 (first entry)
XX	
DE	DNA encoding novel human diagnostic protein #8511.
XX	
XX	
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder; ss

XX	Homo sapiens.
XX	WO200175067-A2.
XX	11-OCT-2001.
XX	30-MAR-2001; 2001WO-US08631.
XX	31-MAR-2000; 2000US-0540217.
XX	23-AUG-2000; 2000US-0649167.
XX	(HYSE-) HYSEQ INC.
XX	Demanac RT, Liu C, Tang YT;
XX	WPI: 2001-639362/73.
XX	P-FSDB; ABG08520.
XX	New isolated polynucleotide and encoded polypeptides, useful in
XX	diagnostics, forensics, gene mapping, identification of mutations
XX	responsible for genetic disorders or other traits and to assess
XX	biodiversity
XX	Claim 1; SEQ ID No 8511; 103pp; English.
XX	The invention relates to isolated polynucleotide (I) and
XX	polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX	and gene mapping, and in recombinant production of (II). The
XX	polynucleotides are also used in diagnostics as expressed sequence tags
XX	for identifying expressed genes. (I) is useful in gene therapy techniques
XX	to restore normal activity of (II) or to treat disease states involving
XX	(II). (II) is useful for generating antibodies against it, detecting or
XX	quantitating a polypeptide in tissue, as molecular weight markers and as
XX	a food supplement. (II) and its binding partners are useful in medical
XX	imaging of sites expressing (II). (I) and (II) are useful for treating
XX	disorders involving aberrant protein expression or biological activity.
XX	The polypeptide and polynucleotide sequences have applications in
XX	diagnostics, forensics, gene mapping, identification of mutations
XX	responsible for genetic disorders or other traits to assess biodiversity
XX	and to produce other types of data and products dependent on DNA and
XX	amino acid sequences. AAS64197-AAS94564 represent novel human
XX	diagnostic coding sequences of the invention.
XX	Note: The sequence data for this patent did not appear in the printed
XX	specification, but was obtained in electronic format directly from WIPO
XX	at ftp.wipo.int/pub/published_pcf_sequences.
XX	Sequence 2121 BP; 515 A; 583 C; 561 G; 462 T; 0 other;
XX	Query Match 63.3%; Score 1064; DB 23; Length 2121;
XX	Best Local Similarity 99.8%; Pred. No. 0;
XX	Matches 1284; Conservative 0; Mismatches 2; Indels 1; Gaps 1.
XX	330 atccctgtctggtgacagcctgtgtgaacacccgcctggaagacacatcatctgtgtgaagtc 389
XX	797 atccctgtgtgtggaagcctgtgtgaacacccgcctggaagacacatcatctgtgtgaagtc 856
XX	390 atccacactaactctctccagctgcaggtgaaagccctgtgccaatgacagagctgtgcta 449
XX	857 atccacactaactctctccagctgcaggtgaaagccctgtgccaatgacagagctgtgcta 916
XX	450 gtcaagatccccctgtgacatgtgtgctgattcaacaacagccctgtgccaagaccactcgtg 509
XX	917 gtcaagatccccctgtgacatgtgtgctgattcaacaacagccctgtgccaagaccactcgtg 976
XX	510 gagttcacatgacgactgaagcccaagccacacatccgcgatgtgacacacagttgaattgac 569
XX	977 gagttcacatgacgactgaagcccaagccacacatccgcgatgtgacacacagttgaattgac 1036
XX	570 cccaccgcctgtgtcctagtgactgtgtccacccgacatgtgagagctgtgcacatcaactg 629



Db 1037 ccacaccgcttggtccctcagtgactgtgcccacagccatggagcctgcgcatccaatcg 1096  
QY 630 ctgcataaagctctctctccttggtgtaagcgcttagcgaagcgtatataagctctcaatg 689  
Db 1097 ctgcataaagctctctcctctggtgaacgccttagcgaagcgtatataagctctcaatg 1156  
QY 690 ccacccctgcggccatctaaatgaanaaacagcgtgtgcccgtatcgaggtcttcctcaat 749  
Db 1157 ccacccctgcggccatctaaatgaanaaacagcgtgtgcccgtatcgaggtcttcctcaat 1216  
QY 750 ggcattatgatcgaacctctcctcagctgtgtgaaggtgcccatttcctcagcatatgacgt 809  
Db 1217 ggcattatgatcgaacctctcctcagctgtgtgaaggtgcccatttcctcagcatatgacgt 1276  
QY 810 ctggaggttgacctctctgctatctctgcatacaaggtgtgaacatttcagctctactggag 869  
Db 1277 ctggaggttgacctctctgctatctctgcatacaaggtgtgaacatttcagctctactggag 1336  
QY 870 gccaaagtgttgactcaagggaaagtgaccgaagtgttcaataactctgacgtctcc 929  
Db 1337 gccaaagtgttgactcaagggaaagtgaccgaagtgttcaataactctgacgtctcc 1396  
QY 930 ctgcataatgccacccctggaacacatcccgctcagccctcatcgtgtagtcagagcgtgtgtg 989  
Db 1397 ctgcataatgccacccctggaacacatcccgctcagccctcatcgtgtagtcagagcgtgtgtg 1456  
QY 990 aaagctgtcagctgtgctgtgtctctcctcagaagaattatgtctgtgttgacctgtgtg 1049  
Db 1457 aaagctgtcagctgtgctgtgtctctcctcagaagaattatgtgtctgtgttgacctgtgtg 1516  
QY 1050 ctctctgagagtgccatcgctcagctgaagtcgaagcatcgagctatcaataaagagctgca 1109  
Db 1517 ctctctgagagtgccatcgctcagctgaagtcgaagcatcgagctatcaataaagagctgca 1576  
QY 1110 gataagctgtgagctctacccagatcgtgaagatcctaactcagaagacatcccgag-ctttt 1168  
Db 1577 gataagctgtgagctctacccagatcgtgaagatcctaactcagaagacatcccgagctttt 1636  
QY 1169 tatagacaaaggcctgcccgaaggtggccaaactgtatcgctggaggtgttccctccag 1228  
Db 1637 tatagacaaaggcctgcccgaaggtggccaaactgtatcgctggaggtgttccctccag 1696  
QY 1228 tgaagccctccgccccttggttcaacctgtgagcatcgaaagcagctcgaaagctcaattta 1288  
Db 1697 tgaagccctccgccccttggttcaacctgtgagcatcgaaagcagctcgaaagctcaattta 1756  
QY 1289 caaccaaaagtgacccaactatactcaacttgaataacatcagctctgtatcgatccagct 1348  
Db 1757 caaccaaaagtgacccaactatactcaacttgaataacatcagctctgtatcgatccagct 1816  
QY 1349 gatgaactctggattgtggtgttccaactgtatgttctgaaanaacatcatactcagagat 1408  
Db 1817 gatgaactctggattgtggtgttccaactgtatgttctgaaanaacatcatactcagagat 1876  
QY 1409 catcacaatcacaatctgtgtgcgaacagaatgtgcaaatataagatctgggtgtccagtgac 1468  
Db 1877 catcacaatcacaatctgtgtgcgaacagaatgtgcaaatataagatctgggtgtccagtgac 1936  
QY 1469 attggtgaagcgctctggagattcgaagcagctgtatcctcactgacgaagaagatgcctgt 1528  
Db 1937 attggtgaagcgctctggagattcgaagcagctgtatcctcactgacgaagaagatgcctgt 1996  
QY 1529 gtttactcagcctctctgttgaaacccagctctcctgtctccagtgaaagcttgagag 1588  
Db 1997 gtttactcagcctctctgttgaaacccagctctcctgtctccagtgaaagcttgagag 2056  
QY 1589 gcagccatcagggaaagctgggtccca 1615  
Db 2057 gcagccatcagggaaagctgggtccca 2083

RESULT 10  
ABA01989

ID ABA01989 standard; cDNA; 1035 BP.  
XX ABA01989;  
AC  
XX  
DT 12-FEB-2002 (first entry)  
XX  
XX Human NOV6b coding sequence.  
DE  
XX  
XX Human; NOV6b; metabolic disorder; neurodegenerative disorder;  
KW immune disorder; haematopoietic disorder; developmental disease; cancer;  
KW retinal disease; feeding disorder; vaccine; infection; gene therapy;  
KW neurological disorder; psychotic disorder; G-protein coupled receptor;  
KW cystatic; antidiabetic; viricide; neuroprotective; nootropic;  
KW analgesic; antidepressant; antimigraine; anticonvulsant; neuroleptic;  
KW antispasmodic; antiallergic; antiinflammatory; anorectic; antiarthritic;  
KW osteopathic; protozoacide; antifungal; hypertensive;  
KW antifertility; vulnery; nephrotoxic; antileptic;  
KW salivary gland protein; chromosome 20; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT 5'UTR 1..79  
FT /\*tag= a  
FT 79..1035  
FT CDS /\*tag= b  
FT /\*product= "NOV6b"  
FT sig\_peptide 79..150  
FT /\*tag= c  
FT mat\_peptide 151..1032  
FT /\*tag= d  
XX  
PN W0200174851-A2.  
XX  
PD 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US10039.  
PF  
XX 30-MAR-2000; 2000US-193205P.  
PR 30-MAR-2000; 2000US-193339P.  
PR 05-APR-2000; 2000US-195343P.  
PR 06-APR-2000; 2000US-195005P.  
PR 06-APR-2000; 2000US-195088P.  
PR 10-APR-2000; 2000US-195792P.  
PR 11-APR-2000; 2000US-196556P.  
PR 13-APR-2000; 2000US-197081P.  
PR 14-APR-2000; 2000US-197087P.  
PR 14-APR-2000; 2000US-197525P.  
PR 29-MAR-2001; 2001US-0823187.  
XX  
XX (CURA-) CURAGEN CORP.  
XX  
XX MaJumder K, Spaderna SK, Taupier RJ, Padigaru M, Burgess CE;  
PI Shinkets RA, Spytek KA, Liu X, Patturajan M, Gusev VY;  
XX WPI: 2001-626379/72.  
XX P-PSDB: AAM47220.  
DR  
XX  
XX New G protein-coupled receptor related polypeptides and polynucleotides  
PT for diagnosis, prevention and treatment of metabolic,  
PT neurodegenerative, retinal, immune, hematopoietic disorders, diabetes,  
PT obesity and infections -  
XX  
XX Claim 8; Page 57; 194pp; English.  
XX  
XX The present invention provides the protein and coding sequences of novel  
CC human G-protein coupled receptors, designated NOV1, NOV2, NOV3, NOV4,  
CC NOV5a, NOV5b, NOV6a, NOV6b, NOV7, NOV8a, NOV8b, NOV9 and NOV10. These can  
CC be used in the treatment of NOVX related diseases, including cancer,  
CC metabolic, neurodegenerative, immune, haematopoietic, developmental,  
CC retinal, feeding, neurological and psychotic diseases and disorders and  
CC infections. The present sequence is the NOV6b coding sequence, the gene



|||||  
Db 302 tgaacacgctccctgaagacacatcatctgctgtaaggtcatcaacagtaacatcctccagc 361  
QY 412 tgaaggtgaagccttgccgaatgaaccagagctgtctatgaagatccccctggaatgg 471  
Db 362 tgaaggtgaagccttgccgaatgaaccagagctgtctatgaagatccccctggaatgg 421  
QY 472 tgcctgattcaacacgcccctgtgtcaagacacatctgtgattccatgaagactgaag 531  
Db 422 tgcctgattcaacacgcccctgtgtcaagacacatctgtgattccatgaagactgaag 481  
QY 532 cccaagccacacatccgcatgagacacagctgcaagtgccaccacgctgtgtcctcagtg 591  
Db 482 cccaagccacacatccgcatgagacacagctgcaagtgccaccacgctgtgtcctcagtg 541  
QY 592 actgtgcacacagcctatggagagctgtgcacatccatctgtcctgaagctcctcctgg 651  
Db 542 actgtgcacacagcctatggagagctgtgcacatccatctgtcctgaagctcctcctgg 601  
QY 652 tgaacgcttagctaaagcagatcatgaacctccttagtgccatcc 695  
Db 602 tgaacgcttagctaaagcagatcatgaacctccttagtgccatcc 645

RESULT 12  
ID AA297023  
AA297023 standard; cDNA; 1175 BP.

AC AA297023;

DT 19-APR-2000 (first entry)

DE Human secreted protein gene 5 cDNA clone HBHMA23, SEQ ID NO:15.

KW Human; secreted protein; cancer; tumour; developmental abnormality;  
KW foetal deficiency; blood disorder; immune system disorder; inflammation;  
KW autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;  
KW schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;  
KW atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;  
KW digestive disorder; endocrine disorder; infection; AIDS; leukaemia;  
KW therapy; ds.

XX Homo sapiens.

XX MO9966041-A1.

XX 23-DEC-1999.

XX 15-JUN-1999; 99WO-US13418.

XX 16-JUN-1998; 98US-0089507.

XX 16-JUN-1998; 98US-0089508.

XX 16-JUN-1998; 98US-0089510.

XX 22-JUN-1998; 98US-0090112.

XX 22-JUN-1998; 98US-0090113.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Ni J, Rosen CA, Wei Y, Young PE, Florence KA;  
PI Soppet DR, Brewer LA, Endress GA, Carter KC, Mucenski M, Ebner R;  
PI Lafleur DW, Olsen HS, Shi Y, Moore PA, Komatsu S G;

XX MPI: 2000-106100/09.

XX P-PSDB: AAY86219.

XX New isolated human genes and the secreted polypeptides they encode,  
PT useful for diagnosis and treatment of e.g. cancers, neurological  
PT disorders, immune diseases, inflammation or blood disorders -

XX Claim 1: Page 318; 586pp: English.

XX AA297019 to AA297137 represent 94 isolated human secreted protein genes.

CC AAY86215 to AAY86333 are the secreted proteins encoded by the 94 human  
CC genes. This sequence represents a fragment of one of the human secreted  
CC proteins. The genes and their corresponding secreted polypeptides are  
CC useful for preventing, treating or ameliorating medical conditions,  
CC e.g., by protein or gene therapy. Also pathological conditions can be  
CC diagnosed by determining the amount of the new polypeptides in a sample  
CC or by determining the presence of mutations in the new genes. Specific  
CC uses are described for each of the 94 genes, based on which tissues they  
CC are most highly expressed in, and include developing products for the  
CC diagnosis or treatment of cancer, tumours, developmental abnormalities  
CC and foetal deficiencies, blood disorders, diseases of the immune system,  
CC autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive  
CC disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin  
CC disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney  
CC disorders, digestive/endocrine disorders, infections and AIDS. The  
CC polypeptides are also useful for identifying their binding partners.  
CC The sequences shown in AAY86334 to AAY86585 represent fragments of the  
CC secreted proteins.

XX SQ Sequence 1175 BP; 290 A; 359 C; 286 G; 239 T; 1 other:

Query Match 31.1%; Score 522; DB 21; Length 1175;  
Best Local Similarity 99.8%; Pred. No. 9,9e-243;  
Matches 642; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 54 gaagcggcggagagactccacgctgcccagctgtgcatcctgactgtgcctctgac 113  
Db 1 gaagcggcggagagactccacgctgcccagctgtgcatcctgactgtgcctctgac 60  
QY 114 accctggaaagtggccgcccgtggacaccttaccctctctgtgttgctggcagccac 173  
Db 61 accctggaaagtggccgcccgtggacaccttaccctctctgtgttgctggcagccac 120  
QY 174 ctgtatccaaagccacctcagctccatcagttcattcattcctctgcccacaaatcatcaa 233  
Db 121 ctgtatccaaagccacctcagctccatcagttcattcattcctctgcccacaaatcatcaa 180  
QY 234 agaaagcgtacacagagagcttgaagagccacacgcccacagatccttgagagctggcc 293  
Db 181 agaaagcgtacacagagagcttgaagagccacacgcccacagatccttgagagctggcc 240  
QY 294 gctcctcagtgccatcgaggaaagccagccgagag-attccctgtgtggcagcctggt 352  
Db 241 gctcctcagtgccatcgaggaaagccagccgagagattccctgtgtggcagcctggt 300  
QY 353 gaacacgctctgaagacacatcattgtgtgaagttcatcaacagcttaacatcctcagct 412  
Db 301 gaacacgctctgaagacacatcattgtgtgaagttcatcaacagcttaacatcctcagct 360  
QY 413 gcaagtgtaagcctctgccaatgaacagagagctgtcagttcaataatcccccttgacatggt 472  
Db 361 gcaagtgtaagcctctgccaatgaacagagagctgtcagttcaataatcccccttgacatggt 420  
QY 473 gcttgatccaacagcccctctgttcaagacacatctgtgagtttccacatgaacgactgagc 532  
Db 421 gcttgatccaacagcccctctgttcaagacacatctgtgagtttccacatgaacgactgagc 480  
QY 533 ccaagccacacatccgcatgagacacagctgaagtggcccccacgctgtgtcctcaagtg 592  
Db 481 ccaagccacacatccgcatgagacacagctgaagtggcccccacgctgtgtcctcaagtg 540  
QY 593 ctgtgcacacagcctatggagagctgtgcacatccatgctgctcaataagctcctcctcctggt 652  
Db 541 ctgtgcacacagcctatggagagctgtgcacatccatgctgctcaataagctcctcctcctggt 600  
QY 653 gaacgcttagctaaagcaggtcatgaacctcctcagtgccatcc 695  
Db 601 gaacgcttagctaaagcaggtcatgaacctcctcagtgccatcc 643

RESULT 13  
AA297114

ID AA297114 standard; cDNA; 1172 BP.  
 AC AA297114.  
 DT 19-APR-2000 (first entry)  
 XX  
 DE Human secreted protein gene 5 cDNA clone HBHMA23, SEQ ID NO:106.  
 XX  
 KW Human; secreted protein; cancer; tumour; developmental abnormality;  
 KW foetal deficiency; blood disorder; immune system disorder; inflammation;  
 KW autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;  
 KW schizophrentia; arthritis; asthma; psoriasis; sepsis; skin disorder;  
 KW atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;  
 KW digestive disorder; endocrine disorder; infection; AIDS; leukaemia;  
 KW therapy; ds.  
 XX  
 OS Homo sapiens.  
 PN W09966041-A1.  
 XX  
 PD 23-DEC-1999.  
 XX  
 PF 15-JUN-1999; 99WO-US13418.  
 XX  
 PR 16-JUN-1998; 98US-0089507.  
 PR 16-JUN-1998; 98US-0089508.  
 PR 16-JUN-1998; 98US-0089509.  
 PR 16-JUN-1998; 98US-0089510.  
 PR 22-JUN-1998; 98US-0090112.  
 PR 22-JUN-1998; 98US-0090113.  
 XX  
 PA (HDMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Ruben SM, Ni J, Rosen CA, Wei Y, Young PE, Florence KA,  
 PI Soppet DR, Brewer LA, Endress GA, Carter KC, Mucenski M, Ebner R;  
 PI Lafleur DW, Olsen HS, Shi Y, Moore PA, Komatsoulis G;  
 XX  
 DR WPI: 2000-106100/09.  
 DR P-PSDB: AAY86310.  
 XX  
 XX New isolated human genes and the secreted polypeptides they encode,  
 PT useful for diagnosis and treatment of e.g. cancers, neurological  
 PT disorders, immune diseases, inflammation or blood disorders -  
 XX  
 PS Claim 1; Page 374; 586pp; English.  
 XX  
 CC AA297019 to AA297137 represent 94 isolated human secreted protein genes.  
 CC AA86215 to AAY86333 are the secreted proteins encoded by the 94 human  
 CC genes. This sequence represents a fragment of one of the human secreted  
 CC proteins. The genes and their corresponding secreted polypeptides are  
 CC useful for preventing, treating or ameliorating medical conditions,  
 CC e.g., by protein or gene therapy. Also pathological conditions can be  
 CC diagnosed by determining the amount of the new polypeptides in a sample  
 CC or by determining the presence of mutations in the new genes. Specific  
 CC uses are described for each of the 94 genes, based on which tissues they  
 CC are most highly expressed in, and include developing products for the  
 CC diagnosis or treatment of cancer, tumours, developmental abnormalities  
 CC and foetal deficiencies, blood disorders, diseases of the immune system,  
 CC autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive  
 CC disorders, schizophrentia, arthritis, asthma, psoriasis, sepsis, skin  
 CC disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney  
 CC disorders, digestive/endocrine disorders, infections and AIDS. The  
 CC polypeptides are also useful for identifying their binding partners.  
 CC The sequences shown in AAY86334 to AAY86585 represent fragments of the  
 CC secreted proteins.  
 XX  
 XX Sequence 1172 BP; 287 A; 358 C; 261 G; 238 T; 8 other;

QY	56	gaggagccgaagatccaagctgcccagctgcttcgatactctgcacatctgtgcctctgacac	115
Db	2	gcgagcccgaggaactcccaagctgcccagctctcgtgacatccctgcacatctgtgcctctgacac	61
QY	116	ctgggaagatgctgcggcccgctggacctcaacctctctctgtgtgttctgtggcaagcaact	175
Db	62	ctgggaagatgctgcggcccgctggacctcaacctctctctgtgtgttctgtggcaagcaact	121
QY	176	tgttccaagccacacctccaatgccactgtcaagtctcatctatcctctggcccaaaagtcataag	235
Db	122	tgtatccaagccacacctccaatgccactgtcaagtctcatctatcctctggcccaaaagtcataag	181
QY	236	aaaagctgtaacaacggagctgtgaaggacccaacaacgcccacagatcctgtgaagatctgcgc	295
Db	182	aaaagctgtaacaacggagctgtgaaggacccaacaacgcccacagatcctgtgaagatctgcgc	241
QY	296	tgtctcaagtgcatactgcgggaaagaacgacgcggagatccctctgtgtgtggcaagctgtgtaa	355
Db	242	tgtctcaagtgcatactgcgggaaagaacgacgcggagatccctctgtgtgtggcaagctgtgtaa	301
QY	356	cacgctccctgaagacacatcatctgttcgaagtgtcatacagaactaatactctcagctgta	415
Db	302	cacgctccctgaagacacatcatctgttcgaagtgtcatacagaactaatactctcagctgta	361
QY	416	ggttgaagcccttcgccaatgtacaagaagctgtctgaatcaagaatctccctctggacatgtgtgc	475
Db	362	ggttgaagcccttcgccaatgtacaagaagctgtctgaatcaagaatctccctctggacatgtgtgc	421
QY	476	tggattccaacaagccctctgctgtctcaagaacatctgtgtgaagtctcaatgaagactgaagccca	535
Db	422	tggattccaacaagccctctgctgtctcaagaacatctgtgtgaagtctcaatgaagactgaagccca	481
QY	536	agccacacatctgcatagtgacacacagtgacagtgcccccacccgctctgtctctgaagtactg	595
Db	482	agccacacatctgcatagtgacacacagtgacagtgcccccacccgctctgtctctgaagtactg	541
QY	596	tcccacacagccatgtggagactgtgcgcatcccaactgtctgtcataaagctcctctctgtgtgaa	655
Db	542	tcccacacagccatgtggagactgtgcgcatcccaactgtctgtcataaagctcctctctgtgtgaa	601
QY	656	cgcccttaagtaagcagtgcatgaacccctcctagtgtgacatcc	695
Db	602	cgcccttaagtaagcagtgcatgaacccctcctagtgtgacatcc	641
RESULT 14			
AAH30554			
ID	AAH30554	standard; cDNA; 420 BP.	
XX	AC	AAH30554;	
DT	27-JUL-2001	(first entry)	
XX	DE	Human colon cancer cell line Km12L4-A cDNA library derived sequence #488	
XX	KW	Human: diagnosis: colon cancer; cancer; malignant; chromosome mapping;	
XX	OS	detection: colon cancer cell line Km12L4-A; ss.	
XX	PN	Homo sapiens.	
XX	PD	MO200018916-A2.	
XX	PE	06-APR-2000.	
XX	PF	23-SEP-1999; 99MO-US22226.	
XX	PR	28-SEP-1998; 98US-0102161.	
XX	PR	28-SEP-1998; 98US-0102180.	
XX	PR	29-SEP-1998; 98US-0102380.	
XX	PR	08-OCT-1998; 98US-0103815.	
XX	PR	27-OCT-1998; 98US-0105877.	
XX	RA	(CHIR ) CHIRON CORP.	

PA (HYSE-) HYSEQ INC.  
 XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;  
 PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;  
 PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;  
 PI Leschovatz D, Kita D, Garcia V, Jones LW, Stache-Crain B;  
 DR WPI; 2000-293155/25.  
 XX Polynucleotide library comprising 1079 defined sequences, useful in  
 PT the form of an array to detect cancer or susceptibility to cancer -  
 XX  
 PS Claim 1: Page 323-324; 502pp; English.  
 CC The present invention describes a library of polynucleotides comprising  
 CC 1079 nucleotide sequences (given in A311067 to A311145). Also described  
 CC are: (1) an isolated polynucleotide (I) having at least 90% identity to  
 CC one of the 1079 sequences; (2) a recombinant host cell containing (1);  
 CC (3) an isolated polypeptide (II) encoded by (1); (4) an antibody that  
 CC specifically binds to (II); (5) a vector comprising (1); and (6) a method  
 CC of detecting differentially expressed genes correlated with a cancerous  
 CC state of a mammalian cell comprising detecting a gene product encoded by  
 CC 65 of the 1079 sequences given in the specification. The polynucleotides  
 CC are used to monitor patients having (or susceptible) to cancer to detect  
 CC potentially malignant events at a molecular level before they are  
 CC detectable at a gross morphological level. The polynucleotides are also  
 CC useful for monitoring the efficacy of various therapies and preventive  
 CC interventions. Polynucleotide probes based on the disclosed sequences  
 CC are useful for chromosome mapping and detection of transcription levels.  
 CC The 1079 polynucleotide sequences were derived from a human colon cancer  
 CC cell line Kml2L4-A cDNA library.  
 XX  
 SQ Sequence 420 BP; 97 A; 116 C; 108 G; 99 T; 0 other;

Query Match 25.0%; Score 420; DB 21; Length 420;  
 Best Local Similarity 100.0%; Pred. No. 2,76-193;  
 Matches 420; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 870 gccaaatttgtagctacagaggaagggagacaaagtgttcaatacctcgaagctcc 929  
 DB 1 gccaaatttgtagctacagaggaagggagacaaagtgttcaatacctcgaagctcc 60  
 QY 930 ctgacatgccacccctgagacacacccgttcagctcattgagtaagagcgtgctg 989  
 DB 61 ctgacatgccacccctgagacacacccgttcagctcattgagtaagagcgtgctg 120  
 QY 990 aagatgctgagctgctgtctctcctcagaagaattcattgctgttgaactctgtg 1049  
 DB 121 aagatgctgagctgctgtctctcctcagaagaattcattgctgttgaactctgtg 180  
 QY 1050 ctctcctgagagtgcccatgctgctgtaagtaagcatggtcgtatcaataaagctgca 1109  
 DB 181 ctctcctgagagtgcccatgctgctgtaagtaagcatggtcgtatcaataaagctgca 240  
 QY 1110 gataagctgagctacacacagatcgtgaagatcctaacacagacactccgaattttt 1169  
 DB 241 gataagctgagctacacacagatcgtgaagatcctaacacagacactccgaattttt 300  
 QY 1170 atagaccagagccatgccaaggttgcccaactgattcgtctggaagtgtttccctcagt 1229  
 DB 301 atagaccagagccatgccaaggttgcccaactgattcgtctggaagtgtttccctcagt 360  
 QY 1230 gaagccctcgcgccttctgttcaacctggtgcatcgaagcagctcgtgaagctcagtttcc 1289  
 DB 361 gaagccctcgcgccttctgttcaacctggtgcatcgaagcagctcgtgaagctcagtttcc 420

RESULT 15  
 ABA01986  
 ID ABA01986 standard; CDNA; 963 BP.  
 XX  
 AC ABA01986;

XX  
 DT 12-FEB-2002 (first entry)  
 XX  
 DE Human NOV6a coding sequence.  
 XX  
 KW Human; NOV6a; metabolic disorder; neurodegenerative disorder;  
 KW immune disorder; hematopoietic disorder; developmental disease; cancer;  
 KW retinal disease; feeding disorder; vaccine; infection; gene therapy;  
 KW neurological disorder; psychotic disorder; G-protein coupled receptor;  
 KW cytostatic; antidiabetic; virucide; neuroprotective; nontropic;  
 KW analgesic; antidepressant; antimigraine; anticonvulsant; neuroleptic;  
 KW antipsoriatic; antiallergic; antiinflammatory; anorectic; antiarthritic;  
 KW osteopathic; protozoacide; antibacter; antihypertensive; hypotensive;  
 KW antinfertility; vulnerary; nephrotropic; antilipemic;  
 KW salivary gland protein; chromosome 20; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FT Key Location/Qualifiers  
 FT CDS 1..963  
 FT /tag= a  
 FT /product= "NOV6a"  
 FT sig\_peptide 1..72  
 FT mat\_peptide 73..960  
 FT /tag= c  
 XX  
 PN WO200174851-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US10039.  
 XX  
 PR 30-MAR-2000; 2000US-193205P.  
 PR 30-MAR-2000; 2000US-193339P.  
 PR 05-APR-2000; 2000US-195343P.  
 PR 06-APR-2000; 2000US-195005P.  
 PR 06-APR-2000; 2000US-195088P.  
 PR 10-APR-2000; 2000US-195792P.  
 PR 11-APR-2000; 2000US-196556P.  
 PR 13-APR-2000; 2000US-197081P.  
 PR 14-APR-2000; 2000US-197087P.  
 PR 14-APR-2000; 2000US-197525P.  
 PR 29-MAR-2001; 2001US-0823187.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 PI Majumder K, Spaderna SK, Taupier RJ, Padigaru M, Burgess CE;  
 PI Shinkets RA, Spytek KA, Liu X, Patturajan M, Gusev VY;  
 PI  
 XX  
 DR WPI; 2001-626379/72.  
 DR P-PSDB; AAM47214.  
 XX  
 PT New G protein-coupled receptor related polypeptides and polynucleotides  
 PT for diagnosis, prevention and treatment of metabolic,  
 PT neurodegenerative, retinal, immune, hematopoietic disorders, diabetes,  
 PT obesity and infections -  
 XX  
 PS Claim 8: Page 54; 194pp; English.  
 XX  
 CC The present invention provides the design and coding sequences of novel  
 CC human G-protein coupled receptors, designated NOV1, NOV2, NOV3, NOV4,  
 CC NOV5a, NOV5b, NOV6a, NOV6b, NOV7, NOV8a, NOV8b, NOV9 and NOV10. These can  
 CC be used in the treatment of NOVX related diseases, including cancer,  
 CC metabolic, neurodegenerative, immune, hematopoietic, developmental,  
 CC retinal, feeding, neurological and psychotic diseases and disorders and  
 CC infections. The present sequence is the NOV6a coding sequence, the gene  
 CC for which is found on chromosome 20. The NOV6a protein shares homology  
 CC with salivary gland proteins.  
 XX  
 SQ Sequence 963 BP; 217 A; 314 C; 237 G; 195 T; 0 other;

Query Match 24.1%; Score 405; DB 22; Length 963;  
 Best Local Similarity 100.0%; Pred. No. 5.1e-186;  
 Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 379 ggctgaaggtatcatcagatcatcctcagctgcaagtgaaagccctcgccaatgacc 438
    |||||||
Db 263 ggctgaaggtatcatcagatcatcctcagctgcaagtgaaagccctcgccaatgacc 322
    |||||||
QY 439 aggaagctctagtcagaatcccccctggaatggtggtggtgttcaacacgcccctgtrca 498
    |||||||
Db 323 aggaagctctagtcagaatcccccctggaatggtggtggtgttcaacacgcccctgtrca 382
    |||||||
QY 499 agaccatcgtggagttccacatgacgactgagggcccaagccacatccgcatggaacacca 558
    |||||||
Db 383 agaccatcgtggagttccacatgacgactgagggcccaagccacatccgcatggaacacca 442
    |||||||
QY 559 gtgcaagtggcccccaccccgctggtcctcaagtgaactgtgccacagcatvggaagctgc 618
    |||||||
Db 443 gtgcaagtggcccccaccccgctggtcctcaagtgaactgtgccacagcatvggaagctgc 502
    |||||||
QY 619 gcatccaactgtgataagctctcctcctcgtggaagccttagctaagcaggtcatga 678
    |||||||
Db 503 gcatccaactgtgataagctctcctcctcgtggaagccttagctaagcaggtcatga 562
    |||||||
QY 679 acctcctagtgccatccctgcccactctagtgaanaaacacagctgtgtcccgltgacgagg 738
    |||||||
Db 563 acctcctagtgccatccctgcccactctagtgaanaaacacagctgtgtcccgltgacgagg 622
    |||||||
QY 739 ctctcctcaatggcatgtatgacagcctcctcgacgctggtgaagg 783
    |||||||
Db 623 ctctcctcaatggcatgtatgacagcctcctcgacgctggtgaagg 667
    |||||||
    
```

Search completed: September 16, 2002, 06:31:11  
 Job time: 7619 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 03:27:07 ; Search time 90.91 Seconds

(without alignments)  
4539.262 Million cell updates/sec

Title: US-09-700-770-2

Perfect score: 1680

Sequence: 1 ggtctgcagatataagtt.....tcaataacactgtcctgtg 1680

Scoring table:

OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PC10S.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfills1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	1.2	7301	4	US-09-816-088-3
2	19	1.1	3471	2	US-08-715-568A-2
3	18	1.1	1406	1	US-08-182-060A-6
4	18	1.1	1406	1	US-08-712-702A-6
5	18	1.1	2068	2	US-08-466-589-1
6	18	1.1	2068	2	US-08-700-636-1
7	18	1.1	2068	3	US-08-467-574-1
8	18	1.1	2068	4	US-09-217-345-1
9	18	1.1	2277	1	US-08-496-855A-1
10	18	1.1	3033	1	US-08-453-695A-111
11	18	1.1	3033	1	US-08-268-161A-111
12	18	1.1	3033	2	US-08-453-702A-111
13	18	1.1	3033	4	US-09-099-639-111
14	18	1.1	3033	5	PCT-US95-08071-111
15	18	1.1	3412	4	US-08-903-325-3
16	18	1.1	11459	4	US-09-454-721A-3
17	17	1.0	864	1	US-08-396-650-2
18	17	1.0	864	1	US-08-768-626-2
19	17	1.0	864	2	US-08-875-062-2
20	17	1.0	921	1	US-08-396-650-3
21	17	1.0	921	1	US-08-396-650-4
22	17	1.0	921	1	US-08-768-626-3
23	17	1.0	921	1	US-08-768-626-4
24	17	1.0	1244	5	PCT-US96-02331-12
25	17	1.0	1355	3	US-08-415-655-14
26	17	1.0	1575	1	US-08-090-523-7
27	17	1.0	1575	1	US-08-398-627-7

28	17	1.0	1575	1	US-08-406-858-7	Sequence 7, Appli
29	17	1.0	1575	5	PCT-US91-04036-7	Sequence 7, Appli
30	17	1.0	1575	5	PCT-US94-05275-7	Sequence 7, Appli
31	17	1.0	2711	4	US-08-891-298-2	Sequence 2, Appli
32	17	1.0	4255	5	PCT-US96-02331-14	Sequence 14, Appli
33	17	1.0	8791	5	PCT-US96-01735-5	Sequence 5, Appli
34	17	1.0	441529	4	US-09-103-840A-1	Sequence 1, Appli
35	16	1.0	27	5	PCT-US94-01263-4	Sequence 1, Appli
36	16	1.0	50	3	US-08-825-852-22	Sequence 22, Appli
37	16	1.0	50	4	US-09-052-888-22	Sequence 22, Appli
38	16	1.0	58	3	US-08-825-852-23	Sequence 23, Appli
39	16	1.0	58	4	US-09-052-888-23	Sequence 23, Appli
40	16	1.0	81	4	US-08-556-978B-27	Sequence 27, Appli
41	16	1.0	81	4	US-08-556-978B-28	Sequence 28, Appli
42	16	1.0	99	4	US-09-199-149-24	Sequence 24, Appli
43	16	1.0	192	2	US-08-442-809A-49	Sequence 49, Appli
44	16	1.0	285	1	US-08-236-520-4	Sequence 4, Appli
45	16	1.0	285	5	PCT-US95-05262-4	Sequence 4, Appli

#### ALIGNMENTS

```

RESULT 1
US-09-816-088-3
; Sequence 3, Application US/09816088
; Patent No. 6326180
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-hui et al
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME, NUCLEIC ACID
; FILE REFERENCE: CLO01181
; CURRENT APPLICATION NUMBER: US/09/816,088
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 7301
; TYPE: DNA
; ORGANISM: Human
US-09-816-088-3

Query Match 1.2% Score 21; DB 4; Length 7301;
Best Local Similarity 100.0%; Pred. No. 0.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 785 gccacatccctcagcatgca 805
Db 3365 gccacatccctcagcatgca 3385

RESULT 2
US-08-715-568A-2
; Sequence 2, Application US/08715568A
; Patent No. 5856463
; GENERAL INFORMATION:
; APPLICANT: Prydz, Hans Peter Blankenborg
; APPLICANT: Brede, Gaute
; TITLE OF INVENTION: PSKH-1 Ribozymes and Uses in Disease
; TITLE OF INVENTION: Treatment
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lerner, David, Littenberg, Krumholz & Mentlik
; STREET: 600 South Avenue West
; CITY: Westfield
; STATE: NJ
; COUNTRY: USA
; ZIP: 07090-1497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/715,568A  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Foley, Shawn P.  
REFERENCE/DOCKET NUMBER: FORSK 3.0-002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 908-654-5000  
TELEFAX: 908-654-7866  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3471 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-715-568A-2

Query Match  
Best Local Similarity 1.1%; Score 19; DB 2; Length 3471;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 134 cgtggaccttcacctct 152  
|||||  
DB 2070 CGTGGACCTTCACCTTCT 2088

RESULT 3  
US-08-182-060A-6  
Sequence 6, Application US/08182060A  
Patent No. 5648210  
GENERAL INFORMATION:  
APPLICANT: John W. Pierce  
APPLICANT: Phillip S. Kerr  
APPLICANT: Mary F. Becker-Manley  
APPLICANT: Richard W. Pearlstein  
APPLICANT: Bruce J. Schweiger  
TITLE OF INVENTION: Nucleotide Sequences  
TITLE OF INVENTION: of Galactinol  
TITLE OF INVENTION: Synthase from  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: E. I. du Pont de Nemours  
ADDRESSEE: and Company  
STREET: 1007 Market Street  
CITY: Wilmington  
STATE: Delaware  
COUNTRY: U.S.A.  
ZIP: 19898  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB  
COMPUTER: Macintosh  
OPERATING SYSTEM: Macintosh System, 6.0  
SOFTWARE: Microsoft Word 4.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/182,060A  
FILING DATE:  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 92/06057  
FILING DATE: 24 JULY 1992  
APPLICATION NUMBER: 07/735,066  
FILING DATE: 24 JULY 1991  
ATTORNEY/AGENT INFORMATION:  
NAME: LINDA AXAMETHY FLOYD  
REGISTRATION NUMBER: 33,692  
REFERENCE/DOCKET NUMBER: BB-1032-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (302)992-4929  
TELEFAX: (302)892-7949

TELEX:  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1406 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA to mRNA  
ORIGINAL SOURCE:  
ORGANISM: Glycine max  
STRAIN: Cultivar Wye  
IMMEDIATE SOURCE:  
LIBRARY: CDNA to mRNA  
CLONE: pS21  
PUBLICATION INFORMATION: Unpublished  
US-08-182-060A-6

Query Match  
Best Local Similarity 1.1%; Score 18; DB 1; Length 1406;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1394 catcatcagagatcat 1411  
|||||  
DB 123 CATCATCCTGAGATCAT 140

RESULT 4  
US-08-712-702A-6  
Sequence 6, Application US/08712702A  
Patent No. 5773699  
GENERAL INFORMATION:  
APPLICANT: John W. Pierce  
APPLICANT: Phillip S. Kerr  
APPLICANT: Mary F. Becker-Manley  
APPLICANT: Richard W. Pearlstein  
APPLICANT: Bruce J. Schweiger  
TITLE OF INVENTION: Nucleotide Sequences  
TITLE OF INVENTION: of Galactinol Synthase from Zucchini and Soybean  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: E. I. du Pont de Nemours  
ADDRESSEE: and Company  
STREET: 1007 Market Street  
CITY: Wilmington  
STATE: Delaware  
COUNTRY: U.S.A.  
ZIP: 18998  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB  
COMPUTER: Macintosh  
OPERATING SYSTEM: Macintosh System, 6.0  
SOFTWARE: Microsoft Word 4.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/712,702A  
FILING DATE: 12-SEP-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/182,060  
FILING DATE: 21-JAN-1994  
APPLICATION NUMBER: US 92/06057  
FILING DATE: 24 JULY 1992  
APPLICATION NUMBER: 07/735,066  
FILING DATE: 24 JULY 1991  
ATTORNEY/AGENT INFORMATION:  
NAME: LINDA AXAMETHY FLOYD  
REGISTRATION NUMBER: 33,692  
REFERENCE/DOCKET NUMBER: BB-1032-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (302)992-4929  
TELEFAX: (302)892-7949  
TELEX:  
INFORMATION FOR SEQ ID NO: 6:



SEQUENCE CHARACTERISTICS:  
LENGTH: 1406 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
ORIGINAL SOURCE:  
ORGANISM: Glycine max  
STRAIN: Cultivar Wye  
IMMEDIATE SOURCE:  
LIBRARY: cDNA to mRNA  
CLONE: PS21  
PUBLICATION INFORMATION: Unpublished  
US-08-712-702A-6

Query Match 1.1%; Score 18; DB 1; Length 1406;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1394 tcacacactagagatcat 1411  
Db 123 CATCATCAGATCATCAT 140

RESULT 5  
US-08-466-589-1  
Sequence 1, Application US/08466589  
Patent No. 5837489  
GENERAL INFORMATION:  
APPLICANT: Elliot, Kathryn J.  
APPLICANT: Ellis, Steven B.  
APPLICANT: Harpold, Michael M.  
TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE  
RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME  
NUMBER OF SEQUENCES: 12  
CURRENT APPLICATION DATA:  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Martin, Haller & McClain  
STREET: 1660 Union Street  
CITY: San Diego  
STATE: CA  
COUNTRY: USA  
ZIP: 92101-2926  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/466,589  
FILING DATE: June 5, 1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/028,031  
FILING DATE: March 8, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie L.  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 6362-9950  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-238-0999  
TELEFAX: 619-238-0062  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2068 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 166..1752

US-08-466-589-1

Query Match 1.1%; Score 18; DB 2; Length 2068;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1396 tcacacactagatcatcc 1413  
Db 1100 TCATCATCAGATCATC 1117

RESULT 6  
US-08-700-636-1  
Sequence 1, Application US/08700636  
Patent No. 5910582  
GENERAL INFORMATION:  
APPLICANT: Elliot, Kathryn J.  
APPLICANT: Ellis, Steven B.  
APPLICANT: Harpold, Michael M.  
TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE  
RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME  
NUMBER OF SEQUENCES: 12  
CURRENT APPLICATION DATA:  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark  
STREET: 444 South Flower Street, Suite 2000  
CITY: Los Angeles  
STATE: CA  
COUNTRY: USA  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/700,636  
FILING DATE: 16-JUL-1996  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/028,031  
FILING DATE: 08-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Reiter, Stephen E.  
REGISTRATION NUMBER: 31,192  
REFERENCE/DOCKET NUMBER: P41 9368  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-546-4737  
TELEFAX: 619-546-9392  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2068 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 166..1752

US-08-700-636-1

Query Match 1.1%; Score 18; DB 2; Length 2068;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1396 tcacacactagatcatcc 1413  
Db 1100 TCATCATCAGATCATC 1117

RESULT 7  
US-08-467-574-1

Sequence 1, Application US/08467574  
Patent No. 6022704  
GENERAL INFORMATION:  
APPLICANT: Elliot, Kathryn J.  
APPLICANT: Ellis, Steven B.  
APPLICANT: Harpold, Michael M.  
TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE  
RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Martin, Haller & McClain  
STREET: 1660 Union Street  
CITY: San Diego  
STATE: CA  
COUNTRY: USA  
ZIP: 92101-2926  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,574  
FILING DATE: June 5, 1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/028,031  
FILING DATE: March 8, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie L.  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 6362-9949  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-238-0999  
TELEFAX: 619-238-0062  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2068 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 166..1752  
US-08-467-574-1

Query Match 1.1%; Score 18; DB 3; Length 2068;  
Best Local Similarity 100.0%; Pred. No. 26;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1396 tcactcactgagatcacc 1413  
DB 1100 TCATCCTGAGATCATCC 1117

RESULT 8  
Sequence 1, Application US/09217345  
Patent No. 6303753  
GENERAL INFORMATION:  
APPLICANT: Elliot, Kathryn J.  
APPLICANT: Ellis, Steven B.  
APPLICANT: Harpold, Michael M.  
TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE  
RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Heller Ehrman White & McCauliffe  
STREET: 4250 Executive Square, 7th Floor  
CITY: La Jolla

STATE: CA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/217,345  
FILING DATE: 21-DEC-98  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/467,574  
FILING DATE: 05-JUN-95  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/466,589,  
FILING DATE: 05-JUN-95  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/028,031  
FILING DATE: 08-MAR-93  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie L.  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 24735-9949B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-450-8400  
TELEFAX: 619-587-5360  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2068 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 166..1752  
US-09-217-345-1

Query Match 1.1%; Score 18; DB 4; Length 2068;  
Best Local Similarity 100.0%; Pred. No. 26;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1396 tcactcactgagatcacc 1413  
DB 1100 TCATCCTGAGATCATCC 1117

RESULT 9  
Sequence 1, Application US/08496855A  
Patent No. 5801232  
GENERAL INFORMATION:  
APPLICANT: Elliot, Kathryn J.  
APPLICANT: Ellis, Steven B.  
APPLICANT: Harpold, Michael M.  
TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE  
RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Martin, Haller & McClain  
STREET: 1660 Union Street  
CITY: San Diego  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92101-2926  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25

;/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/08/496,855A  
/ FILING DATE: 20-JUN-1995  
/ PRIOR APPLICATION DATA:  
/ APPLICATION NUMBER: 08/149,503  
/ FILING DATE: 08-NOV-1993  
/ APPLICATION DATA:  
/ APPLICATION NUMBER: US 08/028,031  
/ FILING DATE: 08-MAR-1993  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Seidman, Stephanie  
/ REGISTRATION NUMBER: 33,779  
/ REFERENCE/DOCKET NUMBER: 6362-9369B  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: 619-238-0999  
/ TELEFAX: 619-238-0062  
/ INFORMATION FOR SEQ ID NO: 1:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 2277 base pairs  
/ TYPE: nucleic acid  
/ STRANDEDNESS: both  
/ TOPOLOGY: both  
/ MOLECULE TYPE: CDNA  
/ FEATURE:  
/ NAME/KEY: CDS  
/ LOCATION: 166..1755  
/ US-08-496-855A-1

Query Match 1.1%; Score 18; DB 1; Length 2277;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1396 tctactgagatgcc 1413  
DB 1103 TCATCAGTGCATCC 1120

RESULT 10  
US-08-453-695A-111  
/ Sequence 111, Application US/08453695A  
/ Patent No. 5708143  
/ GENERAL INFORMATION:  
/ APPLICANT: Suzuki, Shintaro  
/ TITLE OF INVENTION: Protocadherin Materials and Methods  
/ NUMBER OF SEQUENCES: 115  
/ CORRESPONDENCE ADDRESS:  
/ ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
/ STREET: 233 South Wacker, 6300 Sears Tower  
/ CITY: Chicago  
/ STATE: Illinois  
/ COUNTRY: USA  
/ ZIP: 60606  
/ COMPUTER READABLE FORM:  
/ MEDIUM TYPE: Floppy disk  
/ COMPUTER: IBM PC compatible  
/ OPERATING SYSTEM: PC-DOS/MS-DOS  
/ SOFTWARE: Patentin Release #1.0, Version #1.25  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/08/453,695A  
/ FILING DATE:  
/ CLASSIFICATION: 530  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: No. 5708143and, Greta E.  
/ REGISTRATION NUMBER: 35,302  
/ REFERENCE/DOCKET NUMBER: 32658  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: 312/474-6300  
/ TELEFAX: 312/474-0448  
/ TELEX: 25-3856  
/ INFORMATION FOR SEQ ID NO: 111:  
/ SEQUENCE CHARACTERISTICS:

;/ LENGTH: 3033 base pairs  
/ TYPE: nucleic acid  
/ STRANDEDNESS: single  
/ TOPOLOGY: linear  
/ MOLECULE TYPE: CDNA  
/ FEATURE:  
/ NAME/KEY: CDS  
/ LOCATION: 138..2528  
/ US-08-453-695A-111

Query Match 1.1%; Score 18; DB 1; Length 3033;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1051 ttcctgagatgccatc 1068  
DB 562 TTCCTGAGATGCCATC 579

RESULT 11  
US-08-268-161A-111  
/ Sequence 111, Application US/08268161A  
/ Patent No. 5798224  
/ GENERAL INFORMATION:  
/ APPLICANT: Suzuki, Shintaro  
/ TITLE OF INVENTION: Protocadherin Materials and Methods  
/ NUMBER OF SEQUENCES: 115  
/ CORRESPONDENCE ADDRESS:  
/ ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
/ STREET: 233 South Wacker, 6300 Sears Tower  
/ CITY: Chicago  
/ STATE: Illinois  
/ COUNTRY: USA  
/ ZIP: 60606  
/ COMPUTER READABLE FORM:  
/ MEDIUM TYPE: Floppy disk  
/ COMPUTER: IBM PC compatible  
/ OPERATING SYSTEM: PC-DOS/MS-DOS  
/ SOFTWARE: Patentin Release #1.0, Version #1.25  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/08/268,161A  
/ FILING DATE: June 27, 1994  
/ CLASSIFICATION: 435  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Young J. Sun  
/ REGISTRATION NUMBER: P-41,337  
/ REFERENCE/DOCKET NUMBER: 27866/32149  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: 312/474-6300  
/ TELEFAX: 312/474-0448  
/ TELEX: 25-3856  
/ INFORMATION FOR SEQ ID NO: 111:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 3033 base pairs  
/ TYPE: nucleic acid  
/ STRANDEDNESS: single  
/ TOPOLOGY: linear  
/ MOLECULE TYPE: CDNA  
/ FEATURE:  
/ NAME/KEY: CDS  
/ LOCATION: 138..2528  
/ US-08-268-161A-111

Query Match 1.1%; Score 18; DB 1; Length 3033;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1051 ttcctgagatgccatc 1068  
DB 562 TTCCTGAGATGCCATC 579

```
RESULT 12
US-08-453-702A-111
; Sequence 111, Application US/08453702A
; Patent No. 5891706
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: Protocadherin Materials and Methods
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
; STREET: 233 South Wacker, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,702A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5891706and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 32657
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3033 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 138..2528
; US-08-453-702A-111

Query Match
Best Local Similarity 1.1%; Score 18; DB 2; Length 3033;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1051 ttctgtagagtgccatc 1068
|||||
Db 562 TTCCTGAGAGTGCCATC 579

RESULT 13
US-09-099-639-111
; Sequence 111, Application US/09099639
; Patent No. 6262237
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: Protocadherin Materials and Methods
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
; STREET: 233 South Wacker, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/099,639
; FILING DATE: 18 JUN 1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/263,161
; FILING DATE: 27 JUN 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Greta E. No. 6262237and
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/34703
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3033 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 138..2528
; US-09-099-639-111

Query Match
Best Local Similarity 1.1%; Score 18; DB 4; Length 3033;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1051 ttctgtagagtgccatc 1068
|||||
Db 562 TTCCTGAGAGTGCCATC 579

RESULT 14
PCT-US95-08071-111
; Sequence 111, Application PC/TUS9508071
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: Protocadherin Materials and Methods
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/08071
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/12388
; FILING DATE: 23 DEC 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/998,003
; FILING DATE: 29 DEC 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Noland, Greta E.
```

```
;
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 32149
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
;
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3033 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 138..2528
; PCT-US95-08071-111

Query Match 1.1%; Score 18; DB 5; Length 3033;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1051 ttcttgagagtgcccatc 1068
Db 562 TTCCTGAGAGTGCCCATC 579

RESULT 15
US-08-903-325-3
; Sequence 3, Application US/08903325
; Patent No. 6262338
; GENERAL INFORMATION:
; APPLICANT: Peter Schreier; Thomas Herget; and
; TITLE OF INVENTION: RESISTANCE GENES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SPRUNG HORN KRAMER & WOODS
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; MEDIUM TYPE: storage
; COMPUTER: NEC Powermate SX/20
; OPERATING SYSTEM: DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/903,325
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/383,747
; FILING DATE: 27-JAN-1995
; APPLICATION NUMBER: DE 40 31 758.7
; FILING DATE: 06-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/766990
; FILING DATE: 27-SEP-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/235106
; FILING DATE: 28-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Kuit G. Briscoe
; REGISTRATION NUMBER: 33,141
; REFERENCE/DOCKET NUMBER: Bayer 8297.2-KGB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; TELEX:
```

```
;
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3412 nucleotides
; TYPE: Nucleic acid
; STRANDEDNESS: Double
; TOPOLOGY: Linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Peanut (arachis hypogaea)
; US-08-903-325-3

Query Match 1.1%; Score 18; DB 4; Length 3412;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1475 gaagcccttgagattcga 1492
Db 2580 GAAGCCCTTGAGATTGCA 2597
```

Search completed: September 16, 2002, 06:24:15  
Job time: 10628 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 02:58:27 : Search time 3017.73 Seconds  
(without alignments)  
7513.888 Million cell updates/sec

Title: US-09-700-770-2

Perfect score: 1680

Sequence: 1 ggtgtcagagataaagt.....tcataaacactgtcctgtg 1680

Scoring table:

OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estinu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	662	39.4	783	10	BI770190 603053221
2	636	37.9	807	10	BI823956 603039139
3	599	35.7	852	10	BI818205 603032679
4	588	35.0	784	10	BI821790 603035865
5	551	33.8	955	10	BI821769 603035838
6	490	29.2	826	10	BC529820 602558904
7	485	28.9	658	9	BE061388 OVO-BT002
8	451	26.8	567	10	BE088537 RCL-HT088
9	446	26.5	945	10	BI772722 603053235
10	428	25.5	549	10	BC482718 602502690
11	418	24.9	602	10	BC530024 602558870
12	400	23.8	660	10	BC548906 602557379
13	396	23.6	524	10	BE933204 RCL-HT088
14	392	23.3	539	10	BI824243 603040625
15	390	23.2	421	10	BF154920 RCL-BT081
16	386	23.0	465	10	BF154912 RCL-BT081
17	384	22.9	482	10	BF095122 PM3-UT005

18	377	22.4	1026	10	BI821994 603040023
19	374	22.3	469	10	BF743981 RCL-BT081
20	371	22.1	453	9	AI924632 w57e12. x
21	370	22.0	957	10	BI822130 603039831
22	348	20.7	640	10	BC482968 602502981
23	348	20.7	642	10	BF672220 602150403
24	348	20.7	785	10	BC530180 602558678
25	345	20.5	433	10	BF826403 CMA-HN002
26	342	20.4	525	9	AL602621 DKFZP666K
27	336	20.0	438	10	AM167650 xm47a11. x
28	331	19.7	370	10	BE926678 QV2-BT063
29	323	19.2	390	10	BE829540 RCL-ET007
30	322	19.2	365	9	AM381927 OVI-HT031
31	320	19.0	556	10	BE933205 RCL-HT088
32	320	19.0	912	10	BC484717 602505896
33	318	18.9	630	10	BC482764 602502549
34	312	18.6	363	9	AM167610 xm49e05. x
35	305	18.2	379	10	BF088529 RCL-HT088
36	303	18.0	429	9	AM449208 UI-H-B13-
37	296	17.6	482	10	BE926857 OVI-BT067
38	290	17.3	343	9	BE073529 RCL-BT055
39	282	16.8	353	10	BF089104 CMO-HT091
40	277	16.5	743	10	BC548174 602575330
41	273	16.2	836	10	BI820886 603033924
42	264	15.7	397	10	BE693858 RCL-BT055
43	257	15.3	428	10	BE9332793 CMA-BT087
44	246	14.6	318	9	AM996310 QV3-BN004
45	235	14.0	287	10	BF804024 CMA-CT013

#### ALIGNMENTS

RESULT 1  
LOCUS BI770190 783 bp mRNA linear EST 25-SEP-2001  
DEFINITION 603053221F1 NIH\_MGC\_122 Homo sapiens cDNA clone IMAGE:5202893 5',  
mRNA sequence.  
ACCESSION BI770190  
VERSION BI770190.1 GI:15761768  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 783)  
NIH-MGC <http://mgc.ncl.nih.gov/>.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strusberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Life Technologies, Inc.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: LLM11508 row: 1 column: 06  
High quality sequence stop: 782.  
Location/Qualifiers  
1. 783  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5202893"  
/clone\_11b="NIH\_MGC\_122"  
/lab\_host="DH10B"  
/note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6; Site: 1: NotI; Site: 2: EcoRV (destroyed); RNA source anonymous pool of 24 week female lung, 16 week female spleen, and 20-22 week male spleens. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size

#### FEATURES

source

range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 026. Note: this is a NIH-MGC Library."

BASE COUNT 167 a 261 c 199 g 156 t  
ORIGIN

Query Match 39.4%; Score 662; DB 10; Length 783;  
Best Local Similarity 99.9%; Pred. No. 0;

Matches 782; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```

Oy 53 ggaggggcccgaagagctccagcgtgtcccaaggtctgtgcatcctgcaactgtgcccctcta 112
Db 1 ggaggggcccgaagagctccagcgtgtcccaaggtctgtgcatcctgcaactgtgcccctcta 112
Oy 113 cactgggaagatgagccgcccgtggaacttcacacctctctgtgtgtgtgtgtgtgtgtgtgt 172
Db 61 CACCTGGGAAGATGGCCGGCCCGTGGACCTTCACCCCTTCTGTGTGTGTGTGTGTGTGTGTGT 120
Oy 173 ccttgatcaagcaacctcagtcctccactgcaatctcatcctcgtgcccgaagatcatca 232
Db 121 CCTGATCCAAAGCACCCTCAGTCCCACTGCAAGTTCTCATCTCTGGCCCAAAATCATCA 180
Oy 233 aagaaagcttacaagagagctgaagacacaaagcccaagacatcctcagcagctgc 292
Db 181 AAGAAAGCTTACACAGAGAGCTGAAGACCAACAGCCACACAGCATCTCGACAGCTGC 240
Oy 293 cgtgctcagtgcatgcgaggaagagccagcagag -atccctgtctgtgtgtgtgtgtgtgt 351
Db 241 CGCTGCTCAGTGCATCGGGGAAAAAGCCAGCGGAGGATCCCTGTGCTGGGACGCTGG 300
Oy 352 tgaacacgctcctgaagacacatcctgtgtgaagtcatacagactaacatcctccagc 411
Db 301 TGAACACCGTCTTGAACACATCATCTGGCTGAAGTGCATCATCAAGCTAACATCTCTCCAGC 360
Oy 412 tgcagtggaagccctggccaatgacagagagctgtcagtcagatcacccttggaatg 471
Db 361 TGCAGGTGAAGCCCTCGGCCAATGACAGAGAGCTGTGATGATCAAGATCCCTCGGACATGG 420
Oy 472 tgcgtgattacaacgcccctgtgtcaagacacatcgtgtgaagttccacatacagactga 531
Db 421 TGGCTGATTCACACGCGCCCTGGTCAAGACCATCTGTGAGTTCCACATGACGACTGAG 480
Oy 532 cccaagccacatccgatgagacacagctgaagtggtgcccacccgctgtgtcctaagt 591
Db 481 CCCAAGCCACATCCGATGAGACCCAGTGAAGTGGCCCAACCCGCTGTGCTCCTGAG 540
Oy 592 actgtgccaacagacatgagagcctgtgcatacactgtgataagcttcttcttctgt 651
Db 541 ACTGTGCACACAGCATGAGAGCGCTGGCATCACTGCTGCATTAAGCTCTCTTCTG 600
Oy 652 tgaagccttaagtaagcagatcatgaacctctagtcgacatccctgcccacatctagta 711
Db 601 TGAAGCGCTTAAGTAAGCAGTGCATGAACCTCTAGTGCATCCCTGCCAATCTAGTGA 660
Oy 712 aaacacagctgtgtccgtgagatgagagcttctcaatgcatgtatgtgaagactctgc 771
Db 661 AAACACAGCTGTCTCCGTGATGAGGCTTCTCTCAATGACATGATGAGACCTCTCTG 720
Oy 772 agctgtgaagtgcccatctccctcagacatgacagctgtgagttgtgaccttctgtac 831
Db 721 AGCTGTGAAGTGCCCATTTCCCTCAGCATGACCGTGTGAGATTGACCTTCTGTATTC 780
Oy 832 ctg 834
Db 781 CTG 783

```

RESULT 2  
BI823956 807 bp mRNA linear EST 04-OCT-2001  
LOCUS BI823956  
DEFINITION 603039139p1 NIH-MGC\_115 Homo sapiens cDNA clone IMAGE:5180336 5',

ACCESSION mRNA sequence.  
BI823956  
VERSION BI823956.1 GI:15935506  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 807)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgrabs-remail.nih.gov

Tissue Procurement: Life Technologies, Inc.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1AM11449 row: m column: 09  
High quality sequence stop: 793.

FEATURES  
source Location/Qualifiers  
1..807

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5180336"  
/clone\_11b="NIH-MGC\_115"  
/lab\_host="DH10B"  
/note="Organ: pooled brain, lung, testis; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH-MGC Library."

BASE COUNT 175 a 269 c 202 g 161 t  
ORIGIN

Query Match 37.9%; Score 636; DB 10; Length 807;  
Best Local Similarity 99.9%; Pred. No. 6,7e-314;

Matches 756; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```

Oy 62 cgaagactcagcgtgcccagctgtgcatcctgcaactgtgccccttgacactggga 121
Db 7 CGAGGACTCCACCGTCCAGCTGTGCGATCTCTGCTGTGCTGTGCTGTGCTGTGCTGTG 66
Oy 122 agatgcccggcccggtggaacttacccttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 181
Db 67 AGATGGCCGGCCCGTGTGACCTTCTCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 126
Oy 182 aagccacctcagtcctcagctgcagctcctcctcctcctcctcctcctcctcctcctcctc 241
Db 127 AAGCCACCCCTGAGTCCCACTGCAAGTTCATCTCTGCGCCCAAAAGTATCAAAAGAAAGC 186
Oy 242 tgaacagagagctgaaggaacacaaagccacacagatcctgcaagcagctgacctgtca 301
Db 187 TGACACAGAGACTGAAAGACCAACAGCCACACATCTCTGACAGAGTGCCGCTGTGCTCA 246
Oy 302 gtgcacatgcgggaaagacagccggaag -atccctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 360
Db 247 GTGCCATCGGGAAAGCCAGCGAGGACATCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 306
Oy 361 tcttgaagcacatcatctgtgtgaagttcatcaagcctaacaatcctcagctgaggtga 420
Db 307 TCTGGAAGCACATCATCTGTGCTGAAGGTCAACAGCTAACATCTCTCAGCTGAGGTGA 366
Oy 421 agccctcggccaatgaccagagagctgtgtgaagatccctcctgacatgtgtgtgtgtgtgt 480

```



Db	367	AGCCCTGGGCCAATGACAGGAGCTGCTAGTAAATATCCCTGGACATGGTGGCTGAT	426
QY	481	tcacaagccccctggtlcaagaacatcgttggagttccaatgacgatgagcccaagcca	540
Db	427	TCAACACCCCTGGTCAAGACCATCTGGTGAATTCACATGACGATCGAGGCCAAGCCA	486
QY	541	ccatccgatctgacacacagttgcaagtggccccacacgcgtgtgttcctagttagctgtcca	600
Db	487	CCATCCGATGAGACACCATGTCMAATGGCCCCACCCCGCTGGTCTCTCATGTACTGTGCCA	546
QY	601	ccaagccaatgggaagacctgcgcatccaactctgtctcataaagctctcccttcctgtgaaagcct	660
Db	547	CCAGCCATGGGAGGACTCTGGCATCCAACTGCTGCATTAAGTCTCTTCCGTGGTAAGGCT	606
QY	661	tagctaaacaggtcatatgaacctccctcaagtgcattccctgtcccacattagtgaataaacagc	720
Db	607	TAGCTAAACAGGTCAATGAACCTCTCAAGTGCATCTCTGCCAATGTAGTGAATAAACCAAC	666
QY	721	tgtgtcccgatcatgaaggcttccctccaatgtgcattatgacaagacctcctgcagctgtgtga	780
Db	667	TGTGTCCCGTATCGAGGCTTCCTTCATTTGGCATGTATGSCAACAACCTCTCTGCAGCTGTGA	726
QY	781	agggtgccattctccctcaagcatatgaacgctctgtaggtc	817
Db	727	AGGTGCCCATTTCCCTCCAGCATGTACCCCTGTGGAGTT	763

RESULT	3
B1818205	
LOCUS	852 bp mRNA linear EST 04-OCT-2001
DEFINITION	M60303679F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5173842 5', mRNA sequence.
ACCESSION	B1818205
VERSION	B1818205.1 GI:15928527
KEYWORDS	EST.
SOURCE	human.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;  
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.  
1 (bases 1 to 852)  
NIH-MGC <http://mgc.ncl.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.

Tissue Procurement: Life Technologies, Inc.  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLAM11432 row: n column: 19  
 High quality sequence stop: 793.

## FEATURES

### Source

1. .852

```
/organism="Homo sapiens"  
/db_xref="taxon:9606"  
/clone="IMAGE:5173842"  
/clone_1b="NH_MGC_115"  
/lab_host="DH10B"
```

/note="Organ pooled brain, lung, testis; Vector: pCMV-Sport6; Site\_1: NotI; Site\_2: EcoRV (age ranged); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH-MGC Library."

**BASE COUNT**

185 a	282 c	219 g	166 t
-------	-------	-------	-------

25

100

2.

## ORIGIN

Query Match	35.78;	Score 599;	DB 10;	Length 852;
Best Local Similarity	99.98;	Pred. No. 8.4e-295;		
Matches 719;	Conservative 0;	Mismatches 0;	Indels 1;	Gaps 1;

OY	55	agggggccgaggaacccacagctgcccagaagctctgacatccctgacatctgcacctctgaca	114
Db	1	AGCGGGCCGGAGSACTCCAGCGTCGCCAGGTTGGCACTCTGCACTTGTGCGCTCTTACCA	60
OY	115	cctgggaagatgcccggccgctggagccttcaacctctctctgtgtgttctgtgcgcagccacc	174
Db	61	CCTGGGAAGATGGCGCGCGCCCTGGAGCTTCACTTCCTGTGTGTGTGGTGGTGGCAGCCACC	120
OY	175	ttgatccaagccaacctagctccactgtcagttctcatctcttgcgccaaaagtcatcaaa	234
Db	121	TTGATCCCAAGCACCCCTCAGTCCACATGCAGATTCTCATCTCCGGGCCCAAAAGTCATCAAA	180
OY	235	gaaagcctgacacagagactgaagagaccacaacgccaacagcatcctctgcagcaagctgcg	294
Db	161	GAAGAGCTGACACAGAGACTTAAGSACCACACACCACACAGACTCTGCAAGCACTGCGG	240
OY	295	ctgtctcagtgcccatgpcggaagaaacccagcccgagg-attccctgtgtcgtggcagctgtgtg	353
Db	241	CTGCTCATGTGCCATGCGGGGAAAGCCAGCCGAGAGCATCCCTGTGCTGGGACGCTGTGTG	300
OY	354	aacaccgctccctgaagacaatatctctgtgtgaagtcatacagtaatacatcctccagctg	413

Db	301	ATACACCGCTCGAAGCACAATATATCTGGCTGAAGGCTGATCATCAGCTATACATCTCTCCACTG	308
Qy	414	CAAGTGAAGCCCTCGGCAATAGCAAGGAGCTGTCTAGTCAAGATCCTCCCTTGGACATGTG	473
Db	361	CAGGTGAAGCCCTCGGCCAATGACAGGAGCTGTTAGTCAAGATCTCCCTCGACATGTGTG	420
Qy	414	GCTGATATCAACAAGCCCCCTGTCTAAGAATCATGTGTATCTCAATGAAAGCATGTGAAGCC	533
Db	421	GCTGGATTTCAACACGCCCTCTGTGTCAAGACCATGTGGAGTTTCCACATGAGACATGGAAGCC	480
Qy	534	CAAGGCACATCTCGATGTAGACACAGTTCAGTGTGGCCCAACCGACTGTGCTCTCAATGTAC	593

Accession	Sequence	Length
Db	481 CAAGCCACCATCCGATGGACACCGTAGTGGCCCCACCGCGCTGAGTCCATGGAC	540
Qy	594 ttgtccaccagcatatggagccctggcaatccaactgtctgaataagctctcttcctggtg	655
Db	541 TGTGCACACACCAATGGAGGCTGGGCATCCAACTGCTGCATTAAGTCTCTTCCTGGTG	600
Qy	654 aagcccttagctaaagcaggtcatgaacctctagtagcatccctctggcccaatctagtgaaa	713
Db	601 AAGCGCTTAGCTAAGCAGAGTATGAACCTCCTAGTGGCATCCCTGCCCAATCTTAGTAAA	666

[illegible]

ACCESSION	BI821790
VERSION	BI821790.1
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens

Eukaryota; Metazoa; Chordata; Ctenaria; Vertebrata; Euteleostomi  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 784)  
 NIH-MGC <http://mgc.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.



Db 86 AAGCCACCTCAGTCCACTGAGTTCATCTCGGCCCAAAAGTCATCAAAAGAAAGC 145  
Oy 242 tacaagaagagctgaagagaccacaagccaccagcatctctgacgagctgcccgtgtctca 301  
Db 146 TCACACAGAGAGCTGAAGAGCCACAAAGCCACACAGATCTTGCAGACGCTGCGCTGTCTA 205  
Oy 302 gtgcacatgaggaagaaagccagccggaag-attccctgtctggcgagcctgtgtgaacacg 360  
Db 206 GTGCATGCGGGAAAAAGCAGCCGAGGAGCATCTCTGTCTGGGACAGCCTGTGAACACCG 265  
Oy 361 tctgaagaacatcatctgtgctggaagtcatcacagctacaactctccagctgacgtgta 420  
Db 266 TCTTAAAGACATCATCTGTGGCTGAAGGTGATCATCAGACTAAATCTCCGCTGACGAGTGA 325  
Oy 421 agcctcgcgaatgacacagagagctgctgaagatgacatcccccctgtgacatggtgctgagt 480  
Db 326 AGCCCTCGGCCAATACCAAGAGAGCTGCTGATGTAAGATCCCTCGGACATGTTGCTGAT 385  
Oy 481 tcaacagcgccctgtgtcaagaacatcgttgaagttccacaatgacgaactgagcccaagca 540  
Db 386 TCAACAGCGCCCTGTGTCAGACCATGCTGAGTTCACATGACGACTGAGGCCCAAGCCA 445  
Oy 541 ccattcgcatagtgacacacagtgcaagtggccccccagcctgtgtctcgaagtgtgtgca 600  
Db 446 CCATCCGATGAGACACCAATGCAATGGCCCAACCCGCTGCTCTCACTGACTGTGCCA 505  
Oy 601 ccagcattgagagctgtgcatcaccactgtgtcataagctccttcctcctgtgtaacgct 660  
Db 506 CCAGCATGAGAGCTGTGCGCATCCAACTGCTCATAGCTTCTCTTCTGTTGTAACGCTT 565  
Oy 661 tagtaagaagtgatgaacactcctcagtgccatccctggcccaatcagtgtaaaacagc 720  
Db 566 TAGTAAGAGAGTGTATGAACCTCTAGTGCATCCCTGCGCAATGTAGTAAGAAAAACAGC 625  
Oy 721 tgtgtccgtgtgtaagagcttccctcaatgacatgtatgacgaactcctgacgtgtgta 780  
Db 626 TGTGTCCGTGTATGAGGCTTCTCTCAATGAGCATGTATGACAGCTCTGCTGAGTGTGA 685  
Oy 781 aggtgccattt 792  
Db 686 AGTGCCCATTTT 697

RESULT 6  
LOCUS BGS29820 826 bp mRNA linear EST 03-APR-2001  
DEFINITION 6025358904F1 NIH\_MGC\_61 Homo sapiens cDNA clone IMAGE:4696752 5',  
mRNA sequence.  
ACCESSION BGS29820  
VERSION BGS29820.1 GI:13521357  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 826)  
AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LINL at:  
<http://image.llnl.gov>  
Plate: L1CM1524 row: h column: 01  
High quality sequence stop: 820.  
Location/Qualifiers  
1..826  
/organism="Homo sapiens"

FEATURES  
Source

/db\_xref="taxon:9606"  
/clone="IMAGE:4696752"  
/clone.lib="NIH\_MGC.61"  
/tissue\_type="embryonal carcinoma"  
/lab\_host="DH10B (T1 phage resistant)"  
/note="Organ: testis; Vector: pDNR-LIB (Clontech); Site:1:  
SfiI (ggcgccctggcc); Site:2: SfiI (ggccatattggc);  
Double-stranded cDNA was prepared from cell line RNA. 5'  
and 3' adaptors were used in cloning as follows: 5'  
adaptor sequence: 5'-ATTCTAAGGCGGAGGCGCGCATG-3' and 3'  
sequence: 5'-ATTCTAAGGCGGAGGCGCGCATG-3' (30)BN-3'  
(where B = A, C, or G and N = A, C, G, or T). Average  
insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies  
contained inserts by PCR. This library was enriched for  
full-length clones and was constructed by Clontech  
Laboratories (Palo Alto, CA). Note: this is a NIH-MGC  
Library."

BASE COUNT 179 a 271 c 213 g 163 t  
ORIGIN

Query Match 29 2%; Score 490; DB 10; Length 826;  
Best Local Similarity 99.8%; Pred. No. 4,5e-239;  
Matches 610; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Oy 53 ggaagcgagcgagagctccagcgtgccagagctgtgacatcctgcaactgtgccccttga 112  
Db 13 GGAGCGGCGCGAGAGCTCCAGCGTCCCGAGGTCTGGCATCTGCACTTCTGCTCTGTA 72  
Oy 113 caacttgggaagatgagcgccgctgtgacatcaccctctctgtgtgtctgtgcaagca 172  
Db 73 CACGTGGGAAGATGGCGGCGCGGCTGAGACTTCACTTCTGTGTGTTGTGGAGCCA 132  
Oy 173 ccttatccaagccacccctcaatcccatcgtgaagttctatctctgtggcccaaaagtcatca 232  
Db 133 CCTTATCCAAACCCACCTCCAGTCCACATGCAAGTTCTCTATCCCGCCCAAAAGTCATA 192  
Oy 233 aagaaagctgacagagagctgaaaggaacacacagccacagacatcctgtgacagctgc 292  
Db 193 AAGAAAAGCTGACACAGAGAGCTGMAAGACACACAGCCACACGATCTGACGACTGC 252  
Oy 293 cgtgtctcagtgcatctgcggaaagccagccggaag-attccctgtgtctgtggcagccttg 351  
Db 253 CGCTGCTAGTGCATGCGGGAAGCCAGCGGAGGATCCCTGTGCTGGGACAGCTTG 312  
Oy 352 tgaacacgctcctgaagacatcatctgtgtgaaggtcatcaacagctaacatcctccagc 411  
Db 313 TGAACACCGTCTCGAAGCACATCATCTGCTGAAGGTCAATCAAGTCAATCTCCACAC 372  
Oy 412 tgcagtgaaagccctcgcgacatgacacagagagctgctgaagatccccctggagcatg 471  
Db 373 TGCAGGTGAAGCCCTCGGCAATGACACAGAGACTGCTAAGATCCCCCTGGACATGG 432  
Oy 472 tgccttgatccaacacgcccctgtgtcaagaacatcgttgaagttccaacatgacgaactgaag 531  
Db 433 TGGCTGAGATTCAACAGCGCCCTGTGTCAAAGACCATGCTGTGAGTTCCACATGAGAGAG 492  
Oy 532 cccaagccacatctcgcatgtgacacagtgcaagtggcccaaccgctgtgtcctcagtgt 591  
Db 493 CCCAAGCCACCATCCGCAATGACACCAAGTGAAGGGCCGCCCGCTGATCTCAAGG 552  
Oy 592 actgtgcaacagcagcatggaagcctgtgcgcatcacaactgtgtataagctcctcctcgtg 651  
Db 553 ACTGTGCCACGAGCATGGAGAGCTGCGATCCAACTGCTGATAAGCTCTCTTCTGTGG 612  
Oy 652 tgaacgacctta 662  
Db 613 TGAACGCTTTA 623

RESULT 7  
LOCUS BE061388/c BE061388 658 bp mRNA linear EST 09-JUN-2000

DEFINITION QV0-BF02229-251099-038-c01 BF0229 Homo sapiens CDNA, mRNA sequence.  
 ACCESSION BE061388  
 VERSION BE061388.1 GI:8406038  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 658)  
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.  
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 MEDLINE 20202663  
 COMMENT Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?l1=6t2-QV0-BF0229-251099-038-c01&t3=1999-10-25&t4=1>)  
 Seq primer: puc 18 forward  
 High quality sequence stop: 620.  
 Location/Qualifiers  
 1..658  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="BF0229"  
 /dev\_stage="Adult"  
 /note="Organ: breast; Vector: pUC18; Site\_1: Sma1; Site\_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 /716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."  
 BASE COUNT 138 a 161 c 214 g 144 t 1 others  
 ORIGIN  
 Query Match 28.9%; Score 485; DB 9; Length 658;  
 Best local Similarity 99.7%; Pred. No. 1.6e-236;  
 Matches 585; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 656 cgccttagcgaagcaggtatcaactctcagtcacccctccctcccaatctagtgaanaa 715  
 Db 322 CGCCTTAGCCTAAGCAGGTCATGAACCTCTAGGCCATCCCTCCCAATCTAGTGAANA 263  
 QY 716 ccagctgtgcccgtagtcgaagctctctcactgaatgcatgatgagagactctctgaagct 775  
 Db 262 CCAGCTGTGCCCGTAGTCAGAGCTTCTTCAATGGCATGTATGCAACCTCCTGCAGCT 203  
 QY 776 ggtgaaggtgcccatctccctcagcattgacgcgtctggaagtttgaactctgatactgc 835  
 Db 202 GGTGAAGGTGCCCATTTCCCTCAGCATTTGACCTGTGAGATTGACCTCTGTATCTCTGC 143  
 QY 836 catcaaggtggaacacattcactcactcctggggcccaagttgtgactacacaggaa 895  
 Db 142 CATCAAGGTGACACCAATTCATCTTACCTGGGGGCCAAGTTGTGACTCACAGGAAA 83  
 QY 896 ggtgaccagttgttcaataactctgacgtctccctccgaatgcaca 942  
 Db 82 GGTGACCAAGTGTCTCATATACCTCTGCAGCTTCCCTGACAAATGCCA 36  
 RESULT 8  
 BF088537 567 bp mRNA linear EST 19-OCT-2000  
 LOCUS RCI-HT0881-110900-024-d01 HT0881 Homo sapiens CDNA, mRNA sequence.  
 DEFINITION BF088537  
 ACCESSION BF088537.1 GI:10894247  
 VERSION EST.  
 KEYWORDS human.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 567)  
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.  
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 MEDLINE 20202663  
 COMMENT Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?l1=6t2-RC1-HT0881-110900-024-d01&t3=2000-09-11&t4=1>)  
 Seq primer: puc 18 forward  
 High quality sequence start: 19  
 High quality sequence stop: 567.  
 Location/Qualifiers  
 1..567  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="HT0881"  
 /dev\_stage="Adult"  
 /note="Organ: head,neck; Vector: pUC18; Site\_1: Sma1; Site\_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."  
 BASE COUNT 139 a 147 c 143 g 138 t

ORIGIN

Query Match	26.8%;	Score 451;	DB 10;	Length 567;
Best Local Similarity	99.8%;	Pred. No. 3.6e-219;		
Matches 501; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

[illegible]

RESULT	9
LOCUS	BT772722
DEFINITION	BT772722 945 bp mRNA linear EST 25-SEP-2001
ACCESSION	603053235F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5202801 5',
VERSION	BT772722
KEYWORDS	BT772722.1 GI:15764300
SOURCE	EST.
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
JOURNAL	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
COMMENT	1 (bases 1 to 945)
	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .
	National Institutes of Health, Mammalian Gene Collection (MGC)
	Unpublished (1999)
	Contact: Robert Strausberg, Ph.D.
	Email: <a href="mailto:cgabbs-remail.nih.gov">cgabbs-remail.nih.gov</a>
	Tissue Procurement: Life Technologies, Inc.
	CDNA Library Preparation: Life Technologies, Inc.
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
	DNA sequencing by: Incyte Genomics, Inc.
	Clone distribution: MGC clone distribution information can be
	found through the I.M.A.G.E. Consortium/LNLN at:
	<a href="http://image.lnl.gov">http://image.lnl.gov</a>
	plate: LLM1508 row: e column: 10
	High quality sequence stop: 856.
FEATURES	
source	1..945
	Location/Qualifiers

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image:5202801"
/clone_id="NIH_MGC_122"
/lab_host="DH10B"
/notice="Organ: pooled lung and spleen; Vector: pCMV-Sport6;
Site:1: NotI; Site:2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed, and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 026. Note
this is a NIH_MGC Library."

```

Query Match	26.5%;	Score 446;	DB 10;	Length 945;
Best Local Similarity	99.7%;	Pred. No. 1.5e-216;		
Matches 686;	Conservative 0;	Mismatches 0;	Indels 2;	Gaps 2;

Oy	53	ggaagcgagccgaaggaactcaacatcgcccaagttctggaatccctgaacttgcgacctga	112
Db	48	ggagggggcccaaggactctacagtgccccaagstctggcactctgcacttgcgacctga	107
Oy	113	cacctgggaagaatgycgcgcgcctggaaccttaccacctctctctgtgttctctggaacca	117
Db	108	cacctgggaagaatggccggcccggtggaccttcacacctctctgtgtgtgtgtgtggcacca	167
Oy	173	ccctgataccaagccaacctcaagtcctcccaetgaattctcatctctctggcccaaatcatca	233
Db	168	ccctgataccaagccaacctcaagtcctcccaetgaattctcatctctctggcccaaatcatca	227
Oy	233	aagaaagcttacaacaagaagcttgaaggaaccaacaagccacaagcatcctgcaagatgac	29
Db	228	aagaaagcttacaacaagagctgaaaggaacacaaacgacacagcatcctgcaagatgac	287
Oy	293	cgctgtcagtgtccatctgcggaagaagccagcgagga - atccctgtgtctgtgcaagccttg	35
Db	288	cgctgtcagtgtccatctgcggaagaagccagcgagga - atccctgtgtctgtgcaagccttg	347
Oy	352	tgaacacgcgtctctgaagcacatcat - ctgagcttgaagttcatcacagcttaacatccctcag	410
Db	348	tgaacacgcgtctctgaagcacatcatcctgcttgaagttcatcacagcttaacatccctcag	407
Oy	411	ctgcaagtgaagccctcgccaatgacccaaggaagctgtcagttcaagatccccctggacatg	470
Db	408	ctgcaagctgaagccctcgccaatgacccaaggaagctgtcagttcaagatccccctggacatg	467
Oy	471	gtgagctgatatccaagccccctgttcaagaaccatctgtgaggttccaatgaagacgtgag	530
Db	468	gtgagctgatatccaagccccctgttcaagaaccatctgtgaggttccaatgaagacgtgag	527
Oy	531	goccaagccacacatccgaatggaacccaagtgaagaatggccccaccgcgcgtgtccatagt	590
Db	528	goccaagccacacatccgaatggaacccaagtgaagaatggccccaccgcgcgtgtccatagt	587
Oy	591	gaactgtgcacacagccaatgggaagctgtggaatccaactgctgtgatatgaatctctctctg	650
Db	588	gaactgtgcacacagccaatgggaagctgtggaatccaactgctgtgatatgaatctctctctg	647
Oy	651	gtgaacgcctttagctcaagcaggtcaatgaactcttagtgcataccctgcgccaactcattg	710
Db	648	gtgaacgcctttagctcaagcaggtcaatgaactcttagtgcataccctgcgccaactcattg	707
Oy	711	aaaaacagctgtgtccctgtgatacgag 738	
Db	708	aaaaacagctgtgtccctgtgatacgag 735	

## RESULT 10

Query Match	24.98;	Score 418;	DB 10;	Length 602;
Best Local Similarity	99.78;	Pred. No. 2.8e-202;		
Matches 588;	Conservative	0;	Mismatches 1;	Indels 1;
				Gaps 1;

```

QY 53 ggaagcgagccgaagactcgaagctgcccagctcgtgcatccttgccttgcctcctccta 112
    |||||||
Db 13 GGAGCGGGCGGAGAGACTCCAGCGCTGCCAGGCTGGCATCTCAGCTGCTGCCCTGTGA 72

QY 113 caactgggaagatgagcgcgcgtggaacctcaccctctctgtggtttgttgtagcagca 172
    |||||||
Db 73 CACCTGGGAAGATGGCGGGCCGTGGAGCTTCACCCCTTCTCTGTGGTTTCTGGAGCCA 132

QY 173 ccttgatccaagaccaccttgaatcccatgctgagttctcattcctcgcgcgcgaagatca 232
    |||||||
Db 133 CCTTGATCCAAAGCCACCTTCAGTCCCACTGCACTTCATCTCCAGGCCCAAAAGTCATCA 192

QY 233 aagaagaagctgacacaggaagctgaagacacacagccacacagcatccttcagcagctgc 292
    |||||||
Db 193 AAGAAAAGCTGACACAGAGACTGAAGAGCCACACGCCACACATCTTCAGCAGCTGC 252

QY 293 cgcgtgcagtgccatgagcgggaaagccagccggaag-atacctgtgcttggaagccttg 351
    |||||||
Db 253 CGCTGCTCAGTGCCATGGCGGAAAGCCAGCCGAGGCAATCCCTGTGCTGGCAGCCTGG 312

QY 352 tgaacacgcgttcctggaagcaatcattgtgctgaaggttcataagctaaatcctcagc 411
    |||||||
Db 313 TGAACACCGCTCTGAAGCAATCATCTGGCTGAAGGTTCATCAGCTAACATCTCTCAGC 372

QY 412 tgcaggtgaagccctcgcgcacatgacagagctgctagtcaagatcccccttgagatg 471
    |||||||
Db 373 TGAGAGTGAAAGCCCTTGCCCAATGACAGAGCTGCTAGTCAAGATCCCTTGAGACATGG 432

QY 472 tggctgagatccaacagcccttgtgtcaagacacatcgltgaagttccatagacagctgag 531
    |||||||
Db 433 TGCTGATTCACAGCAGCCCTGCTGTCAGACATCGTGAGGTTCCATAGACAGCTGAGG 492

QY 532 cccaagccacatccgcattggaacacagatgagagtgagccacccgcgccttgctcaatg 551
    |||||||
Db 493 CCCAAGCCACATCCCATGAGACACCAAGTGCAGAGTGCACCCCGCTGCTCAGTG 552

QY 592 actgtgccacacgcatlggagagcctgcacatcagctgctgataagctc 641
    |||||||
Db 553 ACTGTGCCACCAACCATGGAGGCTGCGCATTCACACTGCTGCATTAAGCTC 602

RESULT 12
BG548906 660 bp mRNA linear EST 04-APR-2001
LOCUS 602575739F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4703816 5',
DEFINITION mRNA sequence.
ACCESSION BG548906
VERSION BG548906.1 GI:13547571
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 660)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LNCM1542 row: n column: 09
High quality sequence stop: 660.
Location/Qualifiers
1..660
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

```

/clone="IMAGE:4703816"
/clone_lib="NIH_MGC_77"
/lab_host="DH10B (T1 phage-resistant)"
/notes="Organ: Lung; Vector: pDNR-LIB (Clontech); Site:1;
Site1 (ggcgcctcgcgc); Site:2; Site1 (ggcgcattatggc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGCGCATATGAGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGGAGCGGCGCCGATGAG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.9
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH-MGC Library."

BASE COUNT 144 a 222 c 176 g 118 t

ORIGIN

Query Match 23.8%; Score 400; DB 10; Length 660;
Best Local Similarity 99.7%; Pred. No. 4,5e-193;
Matches 570; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 53 ggaagcgagccgaagactcgaagctgcccagctcgtgcatccttgccttgcctcctccta 112
    |||||||
Db 13 GGAGCGGGCGGAGAGACTCCAGCGCTGCCAGGCTGGCATCTCAGCTGCTGCCCTGTGA 72

QY 113 caactgggaagatgagcgcgcgtggaacctcaccctctctgtggtttgttgtagcagca 172
    |||||||
Db 73 CACCTGGGAAGATGGCGGGCCGTGGAGCTTCACCCCTTCTCTGTGGTTTCTGGAGCCA 132

QY 173 ccttgatccaagaccaccttgaatcccatgctgagttctcattcctcgcgcgcgaagatca 232
    |||||||
Db 73 CACCTGGGAAGATGGCGGGCCGTGGAGCTTCACCCCTTCTCTGTGGTTTCTGGAGCCA 132

QY 173 ccttgatccaagaccaccttgaatcccatgctgagttctcattcctcgcgcgcgaagatca 232
    |||||||
Db 133 CTTGATTCACAGCAGCCCTGCTGTCAGACATCGTGAGGTTCCATAGACAGCTGAGG 192

QY 233 aagaagaagctgacacaggaagctgaagacacacagccacacagcatccttcagcagctgc 292
    |||||||
Db 193 AAGAAAAGCTGACACAGAGACTGAAGAGCCACACGCCACACATCTTCAGCAGCTGC 252

QY 293 cgcgtgcagtgccatgagcgggaaagccagccggaag-atacctgtgcttggaagccttg 351
    |||||||
Db 253 CGCTGCTCAGTGCCATGGCGGAAAGCCAGCCGAGGCAATCCCTGTGCTGGCAGCCTGG 312

QY 352 tgaacacgcgttcctggaagcaatcattgtgctgaaggttcataagctaaatcctcagc 411
    |||||||
Db 313 TGAACACCGCTCTGAAGCAATCATCTGGCTGAAGGTTCATCAGCTAACATCTCTCAGC 372

QY 412 tgcaggtgaagccctcgcgcacatgacagagctgctagtcaagatcccccttgagatg 471
    |||||||
Db 373 TGAGAGTGAAAGCCCTTGCCCAATGACAGAGCTGCTAGTCAAGATCCCTTGAGACATGG 432

QY 472 tggctgagatccaacagcccttgtgtcaagacacatcgltgaagttccatagacagctgag 531
    |||||||
Db 433 TGCTGATTCACAGCAGCCCTGCTGTCAGACATCGTGAGGTTCCATAGACAGCTGAGG 492

QY 532 cccaagccacatccgcattggaacacagatgagagtgagccacccgcgccttgctcaatg 551
    |||||||
Db 493 CCCAAGCCACATCCCATGAGACACCAAGTGCAGAGTGCACCCCGCTGCTCAGTG 552

QY 592 actgtgccacacgcatlggagagcctgcacatcagctgctgataagctc 641
    |||||||
Db 553 ACTGTGCCACCAACCATGGAGGCTGCGCATTCACACTGCTGCATTAAGCTC 602

RESULT 13
BE933204 524 bp mRNA linear EST 02-OCT-2000
LOCUS R01-HT0881-290800-022-e03 HT0881 Homo sapiens cDNA, mRNA sequence.
DEFINITION BE933204
ACCESSION BE933204
VERSION BE933204.1 GI:10459280
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

REFERENCE 1 (bases 1 to 524)  
 AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R., Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F., Goldman, G. H., Carvalho, A. F., Matsukuma, A., Baia, G. S., Simpson, D. H., Brunstein, A., deOliveira, P. S., Bucher, P., Joniguel, C. V., O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and Simpson, A. J.  
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 20202663  
 COMMENT Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l=62-rc1-HT0861-290  
 800-022-e03&t3=2000-08-29&t4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 13  
 High quality sequence stop: 524.

FEATURES  
 source  
 1..524  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_1lb="HT0861"  
 /dev\_stage="Adult"  
 /note="Organ: head neck; Vector: puc18; Site:1: SmaI; Site:2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 127 a 145 c 130 g 122 t  
 ORIGIN

Query Match 23.6%; Score 396; DB 10; Length 524;  
 Best Local Similarity 99.8%; Pred. No. 4.8e-191;  
 Matches 446; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 911 caataactctgacgctccctgacaaatgcccacccctgacacacatcccgcttcagcctcat 970  
 DB 78 CAATTAACCTCTCAGCTTCCCTGACAAATGCCACCTGACAAACATCCCGTTCCAGCTCAT 137  
 QY 971 cgttagctgaagctgtggaagctgacgtgacgtgtgctctctccagaagaattcat 1030  
 DB 138 CGTGAGTGAAGACGTGGTGAAGAGCTGCAGTGGCTGTGCTCTCTCCAGAAATTCAT 197  
 QY 1031 gctcctgtgactctgtgctctcctgagagtgcccatcgctggaagtcagcagcgt 1090  
 DB 198 GGTCTCTGTGACCTCTGCTTCCCTGAGAGTGCCTATGCGCTGAAGTCAACATCGGCGCT 257  
 QY 1091 gatacatgaagaagctgagatgaagctgtgattacccaagctcgtggaagatccctaacca 1150  
 DB 258 GATCAATGGAAGGCTGAGATTAAGCTGGATCTAACCCAGATCGTGAAGATCTTAATCA 317  
 QY 1151 ggacactccgagtttttatagaccagagccatcccaagctgagcccaactgactgtct 1210  
 DB 318 GGACACTCCCGAGTTTATATAGACCAAGCCATGCCAAGGAGGCCCAACTGATGCTGCT 377  
 QY 1211 ggaagtgttccctccacagtgaagccctccgaccttgttccacctgggcatcgaaagccag 1270  
 DB 378 GGAAGTGTTCCTCCCTCCAGTGAAGCCCTCCGCGCTTGTTCACCCCGGAGATCGAAGCCAG 437  
 QY 1271 ctgggaagctcagttttacaccaagtgagcaacttatctcaacttgaataaatacag 1330  
 |||||||

DB 438 CTGCGAAGCTCACTTTTACACCAAGGTGACCACTTATTACTGAAATTAATCATGAG 497  
 QY 1331 ctctgacgacccagctgacacac 1357  
 DB 498 CTCTGATCGGATCCAGCTATGATGACTC 524  
 |||||||

RESULT 14  
 LOCUS B1824243 539 bp mRNA linear EST 04-OCT-2001  
 DEFINITION 603040625F1 NIH\_MGC\_115 Homo sapiens cDNA clone IMAGE:5181222 5',  
 mRNA sequence.  
 ACCESSION B1824243  
 VERSION B1824243.1 GI:15935793  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 539)  
 AUTHORS NIH-MGC http://imga.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue procurement: Life Technologies, Inc.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 plate: LLNL1452 row: b column: 07  
 High quality sequence stop: 535.

FEATURES  
 source  
 1..539  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_1lb="IMAGE:5181222"  
 /clone\_1lb="NIH\_MGC\_115"  
 /lab\_host="DH10B"  
 /note="Organ: pooled brain, lung, testis; Vector: pCMV-SPORT6; Site:1: NotI; Site:2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH-MGC Library."

BASE COUNT 119 a 184 c 141 g 94 t  
 ORIGIN

Query Match 23.3%; Score 392; DB 10; Length 539;  
 Best Local Similarity 99.8%; Pred. No. 5.3e-189;  
 Matches 512; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 62 cgaagactccagcgtgcccaagctgacatccctgactgctgcctctgacacactgga 121  
 DB 7 CGAGAGCTCCAGCGCTGCCAGAGCTGTGGCATCTGCTGCTGCTGACCTGACCTGGA 66  
 QY 122 agatggccggccgttggaaccttccacctctgtgtgtgtgtgagcaacctgtatcc 181  
 DB 67 AGATGGCCGGCCGCTGGAGCTTTCACCTTCTGTGTTGCTGGAGCCACCTTGATTC 126  
 QY 182 aagccacctatgctccacatgcaatttcattccctccggcccaaaagtatcaaaagaaagc 241  
 DB 127 AAGCCACCTTATGTTCCACCTGCACTTCATCTCGGCCCCAAAGATATCAAAAGAAAGC 186  
 QY 242 tgacacagagctgaagacacacagccacacacatccctgacacagctggcgtgtctca 301  
 |||||||



```

Db      187 TGACACAGAGAGCTGAAGACACACACAGCAGCATCTGACAGAGCTGCCGTCTCTCA 246
Qy      302 gtgcacatgagggaagaccagccgagag--atccctgtctgagcagccctgtgtaacacg 360
Db      247 GTGCCATGGGGGAAAGCCAGCCGGAGGACATCCCTGTGCTGGGACCTGTGTGAACACCG 306
Qy      361 tccatgaacacatcatctgtctgaaggtatcatcacagctaacatccctcagctcagtgta 420
Db      307 TCCTGAGACACATCATCTGGCTGAAGGTCATCACACCTAACATCCCTCCAGCTCAGAGTCA 366
Qy      421 agccctggagcaatgacagagagctgtctagtaagaatccccctgagacatgtgtgctgat 480
Db      367 ACCCCCTGGGCAATGACCGAGAGCTGCTAGTCAAGATCCCTCGACATGTGTGCTGAT 426
Qy      481 tcaacagccctgtctgaagacacatgttgagttccacatgagcagactgagagcccaagca 540
Db      427 TCAACACGCCCTGTGTCAGAACCATGTGTGAGTCCACATGACGACTGAGGCCCAAGCA 486
Qy      541 ccacccgcatgagacacagtgcaagtgcgccca 573
Db      487 CCATCCGATGACACACCATGCAAGTGGCCCA 519

RESULT 15
BF154920
LOCUS      BF154920                421 bp    mRNA    linear    EST 30-OCT-2000
DEFINITION RCO-BT0812-250900-032-e11 BT0812 Homo sapiens cDNA, mRNA sequence.
ACCESSION  BF154920
VERSION     BF154920.1 GI:11050103
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE   1 (bases 1 to 421)
AUTHORS    Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
            Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
            Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
            20202663
            Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?cl=6r2-RC0-BT0812-250
            900-032-e11&f3=2000-09-25&f4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 3
            High quality sequence stop: 421.
            Location/Qualifiers
                1..421
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_lib="BT0812"
                /dev_stage="Adult"
                /note="Organ: breast; Vector: puc18; Site:1: SmaI; Site:2:
                SmaI; A mini-library was made by cloning products derived
                from ORESTES PCR (U.S. Letters Patent application No. 196
                ,716 - Ludwig Institute for Cancer Research) profiles
                into the pUC 18 vector. Reverse transcription of tissue
                mRNA and cDNA amplification were performed under low
                stringency conditions."
FEATURES
Source

```

```

BASE COUNT      99 a      115 c      107 g      100 t
ORIGIN
Query Match      23.2%; Score 390; DB 10; Length 421;
Best Local Similarity 100.0%; Pred.No. 5.3e-188;
Matches 390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      818 tgacctctgtatccctgcacataaagggtgacccatcagctcactctggtgggccaagt 877
Db      32 TGACCTTGTATCTGTCATCAAGGGTGAACACATTCAGCTGTACTGGGGCCAAAGTT 91
Qy      878 gtggagctacaggggaaggtgacaaagtgttcaactctgacgtccctcgaacat 937
Db      92 GTTGGACTACAGGAAAGGTACCAAGTGTTCATTAATCTGCGAGCTTCCCTGCAAT 151
Qy      938 gcccaacctgagacaacatcccgctcagcctcactgagtcagagcgtggtgaaagctgc 997
Db      152 GCCCACCTTGAGACAAACATCCCGTTACAGCTCATCGTAGTCAAGAGAGTGGTGAAGCTGC 211
Qy      998 agtggctgtgtgtctctcccaagaagaattcaatggtccctgttgactctgtcttcga 1057
Db      212 AGTGGCTGTGTCTCTCTCCAGAGAAATTCATGATGTCGTGTGACTGTGTCTCTCTGA 271
Qy      1058 gaatgccatcggtctgaagtcaagcatcgagctgataatgaaagctgcagataagct 1117
Db      272 GAGTGGCCATCGGCTTAAGTCAGACATCGGGCTGATCATGAAAGCTGCAATAGCT 331
Qy      1118 gggatctaccagatcgtgaaagatccttaactcaagaaactccagagttttatagaca 1177
Db      332 GGGATCTACCCAGATCGTGAAGATCCTTAATCTCAGCACTCCCGAGTTTATAGACA 391
Qy      1178 agggcattgccaaggtgcccactgaatcgt 1207
Db      392 AGGCCATGCCAAGGTGGCCCAACTGATCGT 421

```

Search completed: September 16, 2002, 05:18:12  
Job time: 8385 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 00:52:02 : Search time 390.12 seconds  
(without alignments)  
7393.664 Million cell updates/sec

Title: US-09-700-770-2

Perfect score: 1680

Sequence: 1 ggtgtgcaggtataagtt.....tcaataaaccttgctgtg 1680

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing First 45 summaries

Database :

N.Geneseq.032802:\*

1: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1679.6	100.0	1680	21	AAZ29724
2	1679.2	100.0	1680	22	AAH77948
3	1667	99.2	1707	21	AAH87210
4	1656.2	98.6	2036	22	AAH77950
5	1617	96.2	1636	21	AAA37057
6	1617	96.2	1636	22	AAAS4604
7	1617	96.2	1636	22	AAH92096
8	1617	96.2	1636	22	AAH54293
9	1533	91.2	2121	23	AAH72707

10	806	48.0	1035	22	ABA01989	Human NOV6b coding
11	724.2	43.1	963	22	ABA01986	Human NOV6a coding
12	641	38.2	1172	21	AAZ97114	Human secreted pro
13	635.8	37.8	1177	21	AAH18254	Lung cancer associ
14	634.8	37.8	1175	21	AAZ97023	Human secreted pro
15	420	25.0	420	21	AAH30554	Human colon cancer
16	417.4	24.8	465	21	AAH74959	Human OREF514
17	398.2	23.7	423	22	AAH18484	Human breast cance
18	373.2	22.2	462	22	AAH09812	Human breast cance
19	229	13.6	232	16	AAH21378	Human gene signalu
20	181.4	10.8	594	22	ABA61288	Human foetal liver
21	181.4	10.8	594	22	ABA29112	Probe #7578 for ge
22	181.4	10.8	594	22	AAK09583	Human brain expres
23	181.4	10.8	594	22	AAK35477	Human bone marrow
24	181.4	10.8	594	22	AAH17050	Probe #6983 for ge
25	181.4	10.8	594	22	AAH41191	Probe #9877 used t
26	180	10.7	180	22	ABA3788	Human foetal liver
27	180	10.7	180	22	ABA38963	Probe #17429 for g
28	180	10.7	180	22	AAK22236	Human brain expres
29	180	10.7	180	22	AAK48404	Human bone marrow
30	180	10.7	180	22	AAH26228	Probe #16161 for g
31	180	10.7	180	22	AAH54231	Probe #22917 used
32	163.4	9.7	576	21	AAH74887	Human OREF ORF42
33	161	9.6	583	22	ABA60344	Human foetal liver
34	161	9.6	583	22	ABA28591	Probe #7057 for ge
35	161	9.6	583	22	AAK08624	Human brain expres
36	161	9.6	583	22	AAK34508	Human bone marrow
37	161	9.6	583	22	AAH16752	Probe #6685 for ge
38	161	9.6	583	22	AAH40286	Probe #8912 used t
39	156	9.3	156	22	ABA72883	Human foetal liver
40	156	9.3	156	22	ABA38468	Probe #16934 for g
41	156	9.3	156	22	AAK21316	Human brain expres
42	156	9.3	156	22	AAK47474	Human bone marrow
43	156	9.3	156	22	AAH25947	Probe #15880 for g
44	156	9.3	156	22	AAH53505	Probe #21991 used
45	141	8.4	375	23	AAH72706	DNA encoding novel

ALIGNMENTS

RESULT 1

AAZ29724

AAZ29724 standard; DNA; 1680 BP.

AC AAZ29724:

XX

XX 27-MAR-2000 (first entry)

DT

XX

XX Human Lung specific gene-1.

DE

XX

XX Lung Specific Gene; LSG; human; diagnostic marker;

KW prognosticate; Lung cancer; diagnosis; ds.

KW

XX

XX Homo sapiens.

OS

XX

PN WO960160-A1.

PD

XX

XX 25-NOV-1999.

PF

XX

XX 12-MAY-1999; 99WO-US10344.

PR

XX

XX 21-MAY-1998; 98US-0086212.

XX

XX (DIAD-) DIADEXUS LLC.

XX

XX Yang F, Macina RA, Sun Y;

XX

XX WPI; 2000-116320/10.

DR

XX

XX A new method for diagnosing, monitoring and staging lung cancer

PT

XX

XX Example 1; Pages 33-34; 40pp; English.

PS



FT /transl\_except=" (pos: 328..329, aa: Gly)"  
PT /transl\_except=" (pos: 372..374, aa: Val)"  
XX  
PN  
XX  
PD  
XX  
XX  
PF  
XX  
XX  
PR  
XX  
XX  
PA  
XX  
PI  
XX  
PI  
XX  
DR  
DR  
P-PSDB: AAG63976.

MO200161055-A2.

23-AUG-2001.

20-FEB-2001; 2001WO-US05674.

17-FEB-2000; 2000US-0183188.

(DIAD-) DIADEXUS INC.

Chen S, Sun Y, Macina RA;

WPI: 2001-529917/58.

P-PSDB: AAG63976.

New lung cancer specific gene for the treatment and diagnosis of lung cancer -

Claim 1: Page 112-113; 119pp; English.

XX  
XX  
CC The present sequence represents a human lung cancer specific gene  
CC (Lsg), and encodes a polypeptide designated lng103. Lsgs are useful  
CC in the treatment and diagnosis of lung cancer. The treatment of lung  
CC cancer comprises the administration of a molecule which down regulates  
CC the expression of an Lsg. An immune response can be mounted against a  
CC target cell expressing an Lsg. An identification of potential therapeutic  
CC agents for use in imaging and treating lung cancer which comprises  
CC screening molecules for an ability to bind to or decrease expression  
CC of an Lsg relative to Lsg in the absence of the agent where the ability  
CC of a molecule to bind to the Lsg or decrease expression of the Lsg is  
CC indicative of the molecule being useful in imaging and treating lung  
CC cancer.  
XX  
XX  
SQ

Sequence 1680 BP; 380 A; 503 C; 429 G; 367 T; 1 other;

Query Match

Best Local Similarity 99.9%; Pred. No. 0; DB 22; Length 1680;

Matches 1678; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggtgtcagagataaagtgtgagcttcagaccactgcgcgggagaggagcgagcg 60  
DB 1 ggtgtcagagataaagtgtgagcttcagaccactgcgcgggagaggagcgagcg 60  
QY 61 ccgagagctccagcgtgcccaggtctgagatcctgacttgctgcccctgagacctgg 120  
DB 61 ccgagagctccagcgtgcccaggtctgagatcctgacttgctgcccctgagacctgg 120  
QY 121 aagatgcccggccgtgagccttcacacctctctgtgtgtgtgtgagcagccactgac 180  
DB 121 aagatgcccggccgtgagccttcacacctctctgtgtgtgtgtgagcagccactgac 180  
QY 181 caagcacccctaagtcacacctgagttctcatcctctgcccacaaagtatcaagaaga 240  
DB 181 caagcacccctaagtcacacctgagttctcatcctctgcccacaaagtatcaagaaga 240  
QY 241 ctgagacagagagctgagagacacacagcagcagcatcctgagcagagctgagcgagc 300  
DB 241 ctgagacagagagctgagagacacacagcagcagcatcctgagcagagctgagcgagc 300  
QY 301 agtgcacatcgcgggaaagccagcggagagatccctgtgtgtgtgagcagctgtaacac 360  
DB 301 agtgcacatcgcgggaaagccagcggagagatccctgtgtgtgtgagcagctgtaacac 360  
QY 361 tcttgaagacatcatctgtgtaaggtatcatcagctaacatccctccagctgtgcaagta 420  
DB 361 tcttgaagacatcatctgtgtaaggtatcatcagctaacatccctccagctgtgcaagta 420  
QY 421 agccctcgagcacaatgacagagctgtagtaagaatccctcgagacatgtgtgagat 480  
DB 421 agccctcgagcacaatgacagagctgtagtaagaatccctcgagacatgtgtgagat 480

DB 421 agccctcgagcacaatgacagagctgtagtaagaatccctcgagacatgtgtgagat 480  
QY 481 tcaaacgcccctgtgccaagagacatctgtgagttccacatgacgactgagcccaagcca 540  
DB 481 tcaaacgcccctgtgccaagacacatctgtgagttccacatgacgactgagcccaagcca 540  
QY 541 ccattccgacatgagacagatgcaagtggcccccgcctgtgtctcctcaatgtgagcca 600  
DB 541 ccattccgacatgagacagatgcaagtggcccccgcctgtgtctcctcaatgtgagcca 600  
QY 601 ccaagcactgagagcctgagcatcccaactgctgataaagctctcctcctgtgtaacgct 660  
DB 601 ccaagcactgagagcctgagcatcccaactgctgataaagctctcctcctgtgtaacgct 660  
QY 661 tagctaaagagtgatgaacatctctagtccatccctcgcccacatctagtgaaacacag 720  
DB 661 tagctaaagagtgatgaacatctctagtccatccctcgcccacatctagtgaaacacag 720  
QY 721 tgtgtccgtgatacagagcttctctcaatgagcatgataagacatccctcgagctgtga 780  
DB 721 tgtgtccgtgatacagagcttctctcaatgagcatgataagacatccctcgagctgtga 780  
QY 781 aggtgccatttccctcaagcatgacacgctctgagttgacctctgtatcctgtccatca 840  
DB 781 aggtgccatttccctcaagcatgacacgctctgagttgacctctgtatcctgtccatca 840  
QY 841 aggtgtacacatctcagctctacactgtgagggccaaagtgtgtgtgatacagggaaagtga 900  
DB 841 aggtgtacacatctcagctctacactgtgagggccaaagtgtgtgtgatacagggaaagtga 900  
QY 901 ccaagtggttcaataactctgacagcttccctgtaacatgcccacccggagcaaatccct 960  
DB 901 ccaagtggttcaataactctgacagcttccctgtaacatgcccacccggagcaaatccct 960  
QY 961 tcaagctcatcgtgtgagtcagagacgtgtgtgaaagctgacgtgctgtgtctctccag 1020  
DB 961 tcaagctcatcgtgtgagtcagagacgtgtgtgaaagctgacgtgctgtgtctctccag 1020  
QY 1021 aagaattcagtcctcgtgtgagctgtgtctctcctaagagtgccatctgagtaactaa 1080  
DB 1021 aagaattcagtcctcgtgtgagctgtgtctctcctaagagtgccatctgagtaactaa 1080  
QY 1081 gcatcggcctgatacaatgaaagcgtcgcagataagctggaatcaccagatgtgaaga 1140  
DB 1081 gcatcggcctgatacaatgaaagcgtcgcagataagctggaatcaccagatgtgaaga 1140  
QY 1141 tccatacagagacactcccgagtttttataagacaaagccatgccaaggtggccaaac 1200  
DB 1141 tccatacagagacactcccgagtttttataagacaaagccatgccaaggtggccaaac 1200  
QY 1201 tgaatcgtctgtaagtggttccctccagtgaaagcctccgcttgttcaacctgggca 1260  
DB 1201 tgaatcgtctgtaagtggttccctccagtgaaagcctccgcttgttcaacctgggca 1260  
QY 1261 tgaatcgtctgtaagtggttccctccagtgaaagcctccgcttgttcaacctgggca 1320  
DB 1261 tgaatcgtctgtaagtggttccctccagtgaaagcctccgcttgttcaacctgggca 1320  
QY 1321 ataacatcagctctgatacagatcagctgataagacatggaatgtgctgtgttcaacctg 1380  
DB 1321 ataacatcagctctgatacagatcagctgataagacatggaatgtgctgtgttcaacctg 1380  
QY 1381 atgttctgaaataacatcatcatctgagatcatcactccatcctgtgcccgaacagatg 1440  
DB 1381 atgttctgaaataacatcatcatctgagatcatcactccatcctgtgcccgaacagatg 1440  
QY 1441 gcaaattaagatctggtgtcccgatgcttcatgtgtgaaagccttggaattcgaagagctg 1500  
DB 1441 gcaaattaagatctggtgtcccgatgcttcatgtgtgaaagccttggaattcgaagagctg 1500  
QY 1501 agtccctcactgacaaagatgcccctgtgtcttactcaccagcctcctgtgtgaaacccagct 1560  
DB 1501 agtccctcactgacaaagatgcccctgtgtcttactcaccagcctcctgtgtgaaacccagct 1560

[illegible]

RESULT	3
AAZ87210	
ID	AAZ87210 standard; cDNA; 1707 BP.
XY	

DT	08-MAY-2000	(first entry)
XX		
DE	Human NTAP cDNA clone 2799056.	
XX		

KM Neurotransmission-associated protein; NTP; odorant-binding protein;  
KM neurologic disease; Alzheimer's disease; Parkinson's disease;  
KM Huntington's disease; cerebral neoplasm; multiple sclerosis;  
KM drug screening; gene therapy; antagonist; cancer; AIDS; asthma;  
KM Crohn's disease; osteoporosis; ss.  
KM

05 Homo sapiens.

Key	Location/Qualifiers
FT CDS	124..1578
FT	/*tag= a
FT	/product= "Human neurotrophin-3-associated protein (NTAP) 2799056"

PN WO200001821-A2.

PD 13-JAN-2000.

02-JUL-1999; 99WO-US15121.

PR 02-JUL-1998; 98US-0091677.

PA (INCY-) INCYTE PHARM INC.

PI Lal P, Tang YT, Yue H, Corley NC, Guegler KJ, Gorgone GA;  
PI Baughn MR, Patterson C;

DR WPI; 2000-160770/14.  
DR P-PSDB; AAY77126.

Claim 7, page 67; 67pp; English.

CC Sequences AA287205-287210 represent cDNAs encoding six human  
CC neurotransmission-associated proteins (NTAPs, AA177121-V77126). The  
CC present sequence, clone 2799056, encodes a putative odorant-binding  
CC protein, and was produced by extension of cDNA fragments isolated from a  
CC human nasal polyp tissue cDNA library. The NTAPs are used for treatment  
CC or prevention of neurological diseases (e.g., Alzheimer's, Parkinson's  
CC or Huntington's diseases, cerebral neoplasms, or multiple sclerosis).  
CC They can also be used to raise specific antibodies and to screen for  
CC specific binding agents (potential agonists and antagonists).  
CC NTAP-encoding nucleic acids are useful for recombinant production of  
CC NTAPs, and as a source of therapeutic antagonists (antisenese,  
CC triplex-forming or ribozyme molecules). The nucleic acids may also be  
CC used as a source of probes and primers for diagnosis or monitoring of  
CC NTAP expression in hybridisation/amplification tests, for chromosome  
CC mapping and for identifying related sequences, and for gene therapy. NTAP  
CC antagonists are used to treat and prevent a wide range of cancers and  
CC immune disorders (e.g. AIDS, asthma, Crohn's disease, osteoporosis, viral

CC or other infections). NTAP antibodies are used to detect NTAPs, for  
CC diagnosis or monitoring, as therapeutic antagonists, in competitive drug  
CC screens, and for affinity purification of NTAPs from natural sources.  
XX  
SQ Sequence 1707 BP; 406 A; 503 C; 429 G; 369 T; 0 other;

Query Match	99.2%;	Score 1667;	DB 21;	Length 1707;
Best Local Similarity	99.8%;	Pred. No. 0;		
Matches 1678;	Conservative	1;	Mismatches	1;
			Indels	1;
			Gaps	1;

QY	1	ggctgctgcagagataaaggcttggaacttccagaccacatgcgcccggagagagagggagcggg	60
Db	1	ggctgcagagagataaaggcttggaacttccagaccacacatgcgcccggagagagagagagcggg	60
QY	61	cagagagatccagacgtgcgcccagagctctgcacatccgcgaacttgctctccctctgacaccctggg	120
Db	61	cagagagatccagacgtgcgcccagagctctgcacatccgcgaacttgctctccctctgacaccctggg	120
QY	121	aagatgagccggccgctggaacttcaacccttctctgtgtgtgtgcgagaccacattgac	180
Db	121	aagatgagccggccgctggaacttcaacccttctctgtgtgtgtgcgagaccacattgac	180
QY	181	caagaccaccccacatccccaactgcgaacttctcatctctgcggcccaaaagtcatcaagaagaag	240
Db	181	caagaccaccccacatccccaactgcgaacttctcatctctgcggcccaaaagtcatcaagaagaag	240
QY	241	ctgagcacagagagctggaagagacacaagaacacagagacatctctgcagagagctgcgcgtctc	300
Db	241	ctgagcacagagagctggaagagacacaagaacacagagacatctctgcagagagctgcgcgtctc	300
QY	301	agtgcccatgctcgaggaaaagccaagccagagag-attccctgtgtctgggcagacgtctgtgaaacac	359
Db	301	agtgcccatgctcgaggaaaagccaagccagagagacatccctgtgtctgggcagacgtctgtgaaacac	360
QY	360	gtccctggaagacacatcatctctgctgaaagtgatatacaagcttaacatctctccagagctgcagctg	419
Db	361	gtccctggaagacacatcatctctgctgaaagtgatatacaagcttaacatctctccagagctgcagctg	420
QY	420	aagccctctggccaattgacacagagctgtctatgaataatcccccctgtgacatgctgtgcctgga	479
Db	421	aagccctctggccaattgacacagagctgtctatgaataatcccccctgtgacatgctgtgcctgga	480
QY	480	ttaacaacgagccctctggttcaagaagacacatctgttgaaattccacatgacgagctgagcccaagcc	539
Db	481	ttaacaacgagccctctggttcaagaagacacatctgttgaaattccacatgacgagctgagcccaagcc	540
QY	540	accatccgcacatggacacacagatgacaaatgagctgtgccccacacgagctgtctctcaagtgtatgcc	599
Db	541	accatccgcacatggacacacagatgacaaatgacaaatgagctgtgccccacacgagctgtctctcaagtgtatgcc	600
QY	600	accagacgcataggagagctctgcgacatcaacatgctgcataaagctctctctctgtgtgaagcgc	659
Db	601	accagacgcataggagagctctgcgacatcaacatgctgcataaagctctctctctgtgtgaagcgc	660
QY	660	ttagcttaagcagagatcatgaaacctctcagctcagtgccaactccccaatctaaatgaaataaacag	719
Db	661	ttagcttaagcagagatcatgaaacctctcagctcagtgccaactccccaatctaaatgaaataaacag	720
QY	720	ctgtgtctcccgatgacagagcttcccttcaaatgagcatgtatgacagagaccctctgcagctgtg	779
Db	721	ctgtgtctcccgatgacagagcttcccttcaaatgagcatgtatgacagagaccctctgcagctgtg	780
QY	780	aagagtgaccattccctccaagcatgacgcgtctgagagattgacacctctgatactctgcacac	839
Db	781	aagagtgaccattccctccaagcatgacgcgtctgagagattgacacctctgatactctgcacac	840
QY	840	aagagtgacacacatcaagctctaccctgggggagcagaagtttgatatacagaaggaaagt	899
Db	841	aagagtgacacacatcaagctctaccctgggggagcagaagtttgatatacagaaggaaagt	900
QY	900	accagaatggttcaataactctgcagcttccctgaaatgcccacacctgagacaacatcccg	959

Db 901 accaagtggtcaataactctgcagcttccctcigacaaatgccacccttgacaacatccg 960  
Oy 960 ttcaagccatcgtgtagagacgtggtgaagagctgagctgctgtctcttcca 1019  
Db 961 ttcaagccatcgtgtagagacgtggtgaagagctgagctgctgtctcttcca 1020  
Oy 1020 gaagaaatcagtgctctgtgtagctgtgcttccctgaagatgccatcgctgaagta 1079  
Db 1021 gaagaaatcagtgctctgtgtagctgtgcttccctgaagatgccatcgctgaagta 1080  
Oy 1080 agcatcgagctgatacaatgaagagctgcagataagctggatctacccaagatcgtgaag 1139  
Db 1081 agcatcgagctgatacaatgaagagctgcagataagctggatctacccaagatcgtgaag 1140  
Oy 1140 atccctacacagacactccagatttttatagaccaggccatgcgaagtggcccaa 1199  
Db 1141 atccctacacagacactccagatttttatagaccaggccatgcgaagtggcccaa 1200  
Oy 1200 ctgatactgctggaagtgttccctccagtgaaagccctccgacctgttccacctgggc 1259  
Db 1201 ctgatactgctggaagtgttccctccagtgaaagccctccgacctgttccacctgggc 1260  
Oy 1260 atcgaagcagactcggaagctcaagttttacaccaaaagtgcaccaactataactgaactg 1319  
Db 1261 atcgaagcagactcggaagctcaagttttacaccaaaagtgcaccaactataactgaactg 1320  
Oy 1320 aataaataagctctgatactgagatccagctgataactctggagttgctggttccaaact 1379  
Db 1321 aataaataagctctgatactgagatccagctgataactctggagttgctggttccaaact 1380  
Oy 1380 gatgtctgaaaaacatcatcactgtagatcatccatcatcctgtgtccgaacagaat 1439  
Db 1381 gatgtctgaaaaacatcatcactgtagatcatccatcatcctgtgtccgaacagaat 1440  
Oy 1440 ggaacaataagatctgggtcccaatgtagatctgtagaagccttggagatcgagagact 1499  
Db 1441 ggaacaataagatctgggtcccaatgtagatctgtagaagccttggagatcgagagact 1500  
Oy 1500 gagtcctcactgacccaagagatgacctgtgtctactccagacctctgtggaaccagc 1559  
Db 1501 gagtcctcactgacccaagagatgacctgtgtctactccagacctctgtggaaccagc 1560  
Oy 1560 tctccctgtcccaagtagaacttgatgtagcagccaatcagggaagctgggtccagctg 1619  
Db 1561 tctccctgtcccaagtagaacttgatgtagcagccaatcagggaagctgggtccagctg 1620  
Oy 1620 ggaagtagtggtgtagctctatagaccatccctctctcgaatcaataaacaacttgctgt 1679  
Db 1621 ggaagtagtggtgtagctctatagaccatccctctctcgaatcaataaacaacttgctgt 1680  
Oy 1680 g 1680  
Db 1681 g 1681

RESULT 4  
AAH77950  
ID AAH77950 standard; DNA; 2036 BP.  
XX  
AC AAH77950;  
XX  
DT 13-NOV-2001 (first entry)  
XX  
DE Nucleotide sequence of a human Lng103 polypeptide.  
XX  
KW Human; lung cancer specific gene; LSG; Lng103; lung cancer; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 163..1617  
FT /tag= a  
FT /product= "Lng103"

FT  
XX  
PN WO200161055-A2. /transl\_except= "(pos: 412..414, aa: Val)"  
XX  
PD 23-AUG-2001.  
XX  
XX  
PF 20-FEB-2001; 2001WO-US05674.  
XX  
PR 17-FEB-2000; 2000US-0183188.  
XX  
PA (DIAD-) DIADEXUS INC.  
XX  
PI Chen S, Sun Y, Macina RA;  
XX  
DR WPI: 2001-529917/58.  
DR P-PSDB: AAG63976.  
PT  
PT New lung cancer specific gene for the treatment and diagnosis of lung  
XX cancer -  
PS Claim 1: Page 116-117; 119pp; English.  
XX  
CC The present sequence represents a human lung cancer specific gene  
CC (LSG), and encodes a polypeptide designated Lng103. LSGs are useful  
CC in the treatment and diagnosis of lung cancer. The treatment of lung  
CC cancer comprises the administration of a molecule which down regulates  
CC the expression of an LSG. An immune response can be mounted against a  
CC target cell expressing an LSG. Identification of potential therapeutic  
CC agents for use in imaging and treating lung cancer which comprises  
CC screening molecules for an ability to bind to or decrease expression  
CC of an LSG relative to LSG in the absence of the agent where the ability  
CC of a molecule to bind to the LSG or decrease expression of the LSG is  
CC indicative of the molecule being useful in imaging and treating lung  
CC cancer.  
XX  
SQ Sequence 2036 BP; 477 A; 586 C; 528 G; 444 T; 1 other:  
XX  
Query Match 98.6%; Score 1656.2; DB 22; Length 2036;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1678; Conservative 2; Mismatches 0; Indels 3; Gaps 2;  
Oy 1 ggtgtcagagataaagattgagacttccagaccactgcccggagagc 58  
Db 38 ggtgtcagagataaagattgagacttccagaccactgcccggagagagagagc 97  
Oy 59 ggcgcagagactcagagctgcccagatctgagatccctgactgtgcccctgtacacctg 118  
Db 98 ggcgcagagactcagagctgcccagatctgagatccctgactgtgcccctgtacacctg 157  
Oy 119 ggaagatgagccggccgctgagacttcaacctctctgtgtgttgcggagccacttga 178  
Db 158 ggaagatgagccggccgctgagacttcaacctctctgtgtgttgcggagccacttga 217  
Oy 179 tccaagccacctcagtgcccaactgagttctatccctcgcccaaaagtatcaaaagaa 238  
Db 218 tccaagccacctcagtgcccaactgagttctatccctcgcccaaaagtatcaaaagaa 277  
Oy 239 agctgacacagagagctgaagagacacaacgcacacagatcctgtcagcagctgcgctgc 298  
Db 278 agctgacacagagagctgaagagacacaacgcacacagatcctgtcagcagctgcgctgc 337  
Oy 299 tcaagtgcacatgcgggaaagccagccggagg -atccctgtgctgggagcagctgtgaaca 357  
Db 338 tcaagtgcacatgcgggaaagccagccggaggatccctgtgctgggagcagctgtgaaca 397  
Oy 358 ccgtctgaaagacatcatctgtgctgaagtatcacaagtaaacatccctccaactcagc 417  
Db 398 ccgtctgaaagacatcatctgtgctgaagtatcacaagtaaacatccctccaactcagc 457  
Oy 418 tgaagccctcgagcaatgagcagagctgtagtcaagaatcccccctgagatgtgtgctg 477  
Db 458 tgaagccctcgagcaatgagcagagctgtagtcaagaatcccccctgagatgtgtgctg 517

QY	478	gatlcaaacagccccgtgtgtcaagaacacatctgtgtgagttccacaatgacagctgtgagccccag	537
Db	518	gatlcaaacacagccccgtgtgtcaagaacacatctgtgtgagttccacaatgacagctgtgagccccag	577
QY	538	ccacatccgcgcgtgagcacacagctgtgacagtgtgccccacccccgcctgtgtccatgtactgtg	597
Db	578	ccacacacccgcgcgtgagcacacagctgtgacagtgtgccccacccccgcctgtgtccatgtactgtg	637
QY	598	ccacacagcccatgttggaagccgtgcgcgcatcccaactctgtgtcaaaagctctccctccctgttgag	657
Db	638	ccacacagcccatgttggaagccgtgcgcgcatcccaactctgtgtcaaaagctctccctccctgttgag	697
QY	658	cccttaagctaaagcgcgtgtcaatgaaacctccctagtgccatccctgtgccccaatctagtgtgaaaac	717
Db	698	cccttaagctaaagcgcgtgtcaatgaaacctccctagtgccatccctgtgccccaatctagtgtgaaaac	757
QY	718	agctgtgtccccgtgtactgcagagctctccctcaatgagcatgtatgcagagactctctgcagctg	777
Db	758	agctgtgtccccgtgtactgcagagctctccctcaatgagcatgtatgcagagactctctgcagctg	817
QY	778	tgaaggtgcacatttccctccagacattgacgcgtgtgagattgtgacctgtatccctgtca	837
Db	818	tgaaggtgcacatttccctccagacattgacgcgtgtgagattgtgacctgtatccctgtca	877
QY	838	tcaaggtgtacacacattacagctctacacctgtgggggcacaagttgttgactccaaaggaaaag	897
Db	878	tcaaggtgtacacacattacagctctacacctgtgggggcacaagttgttgactccaaaggaaaag	937
QY	898	tgaaccaagtgtgtcaataactctctgcagctctccctgtacacatgtccccacctgtgaaacacatcc	957
Db	938	tgaaccaagtgtgtcaataactctctgcagctctccctgtacacatgtccccacctgtgaaacacatcc	997
QY	958	cgltcagccctcaatcgttgagtcagagcagctgtgtgaaagctgtgaagctgtgcctgtgtctctcc	1017
Db	998	cgltcagccctcaatcgttgagtcagagcagctgtgtgaaagctgtgaagctgtgcctgtgtctctcc	1057
QY	1018	cagaagaattcatgtgtccctgtgtgagctctgtgctctccctgtgagagttgcccacatcgctgtgaagt	1077
Db	1058	cagaagaattcatgtgtccctgtgtgagctctgtgctctccctgtgagagttgcccacatcgctgtgaagt	1117
QY	1078	caagcatcggggtgtgtcaatgtaaagaagctgtcaagatagcgtgggacttaccacagactcgtgga	1137
Db	1118	caagcatcggggtgtgtcaatgtaaagaagctgtcaagatagcgtgggacttaccacagactcgtgga	1177
QY	1138	agatctcaatcaccagagacacctccagagtttttataagaccagaagcccatgtccaaagtgtgccc	1197
Db	1178	agatctcaatcaccagagacacctccagagtttttataagaccagaagcccatgtccaaagtgtgccc	1237
QY	1198	aactgtactgtgctgtgaaggtgttctccctccaggtgaagccccctccgaccttgtgttccacctg	1257
Db	1238	aactgtactgtgctgtgaaggtgttctccctccaggtgaagccccctccgaccttgtgttccacctg	1297
QY	1258	gcactgtgaagccagctcgggaagctcagtttataaccacaaagtgtagccaactataactcaact	1317
Db	1298	gcactgtgaagccagctcgggaagctcagtttataaccacaaagtgtagccaactataactcaact	1357
QY	1318	tgaataaactaagctctgtcatcgcgactcgaactgtgataactctgtgagattgtgctgtgttccaac	1377
Db	1358	tgaataaactaagctctgtcatcgcgactcgaactgtgataactctgtgagattgtgctgtgttccaac	1417
QY	1378	ctgtatgttctgtaaaaaacatcatcatcactgtgagatcaatccactccactctgtctgcgcgaacaga	1437
Db	1418	ctgtatgttctgtaaaaaacatcatcatcactgtgagatcaatccactccactctgtctgcgcgaacaga	1477
QY	1438	atgtgcaaatlaagactctgtgggtctccagtgatcaattgtgtgaaagcctctgtgattctgagagag	1497
Db	1478	atgtgcaaatlaagactctgtgggtctccagtgatcaattgtgtgaaagcctctgtgattctgagagag	1537
QY	1498	ctgtgctctcaactgtgacccaagaagctgtgtgttacttaccacagctcctctgtgtgaaaccca	1557
Db	1538	ctgtgctctcaactgtgacccaagaagctgtgtgttacttaccacagctcctctgtgtgaaaccca	1597

Qy	1558	gctctcccttctccccaagtgaagacttggaatggcagcatcagggaagactggctccacg	1617
Db	1598	gctctcccttctccccaagtgaagacttggaatggcagcatcagggaagactggctccacg	1657
Qy	1618	tgaggatcaggtgtgagctctatagaccatccctctctgcaataaacttgcct	1677
Db	1658	tgaggatcaggtgtgagctctatagaccatccctctctgcaataaacttgcct	1717
Qy	1678	gtg 1680	
Db	1718	gtg 1720	
RESULT 5			
AAA37057			
ID	AAA37057 standard; cDNA: 1636 BP.		
XX	AAA37057;		
AC			
XX			
DT	08-AUG-2000 (first entry)		
XX			
DE	Human PRO1357 (UNQ706) cDNA sequence SEQ ID NO:127.		
XX			
KM	Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;		
KW	transmembrane; secretion; immunoadhesion; pharmaceutical; screening;		
XX	ss.		
OS	Homo sapiens.		
XX			
PN	WO200012708-A2.		
XX			
PD	09-MAR-2000.		
XX			
PF	01-SEP-1999; 99MO-US20111.		
XX			
PR	01-SEP-1998;	98US-0098716;	
PR	01-SEP-1998;	98US-0098749;	
PR	01-SEP-1998;	98US-0098750;	
PR	02-SEP-1998;	98US-0098803;	
PR	02-SEP-1998;	98US-0098821;	
PR	02-SEP-1998;	98US-0098843;	
PR	09-SEP-1998;	98US-0099536;	
PR	09-SEP-1998;	98US-0099596;	
PR	09-SEP-1998;	98US-0099602;	
PR	09-SEP-1998;	98US-0099642;	
PR	10-SEP-1998;	98US-0099741;	
PR	10-SEP-1998;	98US-0099754;	
PR	10-SEP-1998;	98US-0099763;	
PR	10-SEP-1998;	98US-0099792;	
PR	10-SEP-1998;	98US-0099808;	
PR	10-SEP-1998;	98US-0099812;	
PR	10-SEP-1998;	98US-0099815;	
PR	15-SEP-1998;	98US-0099816;	
PR	15-SEP-1998;	98US-0100385;	
PR	15-SEP-1998;	98US-0100388;	
PR	15-SEP-1998;	98US-0100390;	
PR	16-SEP-1998;	98US-0100584;	
PR	16-SEP-1998;	98US-0100627;	
PR	16-SEP-1998;	98US-0100661;	
PR	16-SEP-1998;	98US-0100662;	
PR	17-SEP-1998;	98US-0100664;	
PR	17-SEP-1998;	98US-0100683;	
PR	17-SEP-1998;	98US-0100684;	
PR	17-SEP-1998;	98US-0100710;	
PR	17-SEP-1998;	98US-0100711;	
PR	17-SEP-1998;	98US-0100913;	
PR	18-SEP-1998;	98US-0100930;	
PR	18-SEP-1998;	98US-0100848;	
PR	18-SEP-1998;	98US-0100849;	
PR	18-SEP-1998;	98US-0101014;	
PR	18-SEP-1998;	98US-0101066;	
PR	18-SEP-1998;	98US-0101071;	





QY 530 ggcccaagccacatccgcgcatggacaccagtggaagtgccccacccgcctgctctccag 589  
|||||  
Db 481 ggcccaagccacatccgcgcatggacaccagtggaagtgccccacccgcctgctctccag 540  
QY 590 tgactgtgcaaccgacatgggaagctgcgcatccaaactgtctgataaagctctctctc 649  
|||||  
Db 541 tgactgtgcaaccgacatgggaagctgcgcatccaaactgtctgataaagctctctctc 600  
QY 650 ggtgaacgcttaagctaaagctgcatgaactctctctagtgccatccctggcccaatctagt 709  
|||||  
Db 601 ggtgaacgcttaagctaaagctgcatgaactctctctagtgccatccctggcccaatctagt 660  
QY 710 gaaacaacagctgtgtcccgtagtcgaaggtctctctcaatgagcatgfatgagacctctc 769  
|||||  
Db 661 gaaacaacagctgtgtcccgtagtcgaaggtctctctcaatgagcatgfatgagacctctc 720  
QY 770 gcagctgtggaagtgccccatttccctcagatgacccgtctgaggtttgaacctctgta 829  
|||||  
Db 721 gcagctgtggaagtgccccatttccctcagatgacccgtctgaggtttgaacctctgta 780  
QY 830 tccctgcatcaaggggtgacacattctacgtctgggggccaagttgtttgagctaca 889  
|||||  
Db 781 tccctgcatcaaggggtgacacattctacgtctgggggccaagttgtttgagctaca 840  
QY 890 ggggaaggtgacccaagtgtgtcaataactctgcaagcttccctgacaatgtgccacctgga 949  
|||||  
Db 841 ggggaaggtgacccaagtgtgtcaataactctgcaagcttccctgacaatgtgccacctgga 900  
QY 950 caaatcccgcttcagctcatctgtgagtcaggaagctgtgaaagctgcaatgtgctgtgt 1009  
|||||  
Db 901 caaatcccgcttcagctcatctgtgagtcaggaagctgtgaaagctgcaatgtgctgtgt 960  
QY 1010 gctctctcagaagaattcatggtctctgtgtgactctgtctctctcagagtgccccatg 1069  
|||||  
Db 961 gctctctcagaagaattcatggtctctgtgtgactctgtctctctcagagtgccccatg 1020  
QY 1070 gctgaagctcagatcggtggtgtcgaatgaaaggtctgcagataagctggagctctacca 1129  
|||||  
Db 1021 gctgaagctcagatcggtggtgtcgaatgaaaggtctgcagataagctggagctctacca 1080  
QY 1130 gatcgtgaaagatcctactacagacactcccgagtttttatagacaaagggcatcccaa 1189  
|||||  
Db 1081 gatcgtgaaagatcctactacagacactcccgagtttttatagacaaagggcatcccaa 1140  
QY 1190 ggtgtgcccaactgatacgtgtcgtgaagtggttccctccagtgaaagccctccgacctgtgt 1249  
|||||  
Db 1141 ggtgtgcccaactgatacgtgtcgtgaagtggttccctccagtgaaagccctccgacctgtgt 1200  
QY 1250 caccctgggcatcgaaagcagctcggaaagctcaagttttacaccaaaagtgacaaacttat 1309  
|||||  
Db 1201 caccctgggcatcgaaagcagctcggaaagctcaagttttacaccaaaagtgacaaacttat 1260  
QY 1310 actcaacttgaaataacatagctctctatcgatccagctgataagctctgggattggctg 1369  
|||||  
Db 1261 actcaacttgaaataacatagctctctatcgatccagctgataagctctgggattggctg 1320  
QY 1370 gttccaacttgatgttctgtaaaaaatacatcaactagatcatcatccatccatctctgtgc 1429  
|||||  
Db 1321 gttccaacttgatgttctgtaaaaaatacatcaactagatcatccatccatccatctctgtgc 1380  
QY 1430 gaacccaagtgcacaatgaatgactcgggggtcccaagtgatcagtggaagggcttggatt 1489  
|||||  
Db 1381 gaacccaagtgcacaatgaatgactcgggggtcccaagtgatcagtggaagggcttggatt 1440  
QY 1490 cgaagcagctgagtcctcaactgaccaaagatgaccttgctgacttaaccagctcctgtgtg 1549  
|||||  
Db 1441 cgaagcagctgagtcctcaactgaccaaagatgaccttgctgacttaaccagctcctgtgtg 1500  
QY 1550 gaaacccaagctctcgtctccagtgaaagattgtagtgagcgacatcaagggaaggtctgg 1609  
|||||  
Db 1501 gaaacccaagctctcgtctccagtgaaagattgtagtgagcgacatcaagggaaggtctgg 1560

QY 1610 gtcccaagctggaggtatgggtgtgagctctatagacatccctctctgcaataaataaac 1669  
|||||  
Db 1561 gtcccaagctggaggtatgggtgtgagctctatagacatccctctctgcaataaataaac 1620  
QY 1670 acttgctctg 1680  
|||||  
Db 1621 acttgctctg 1631  
RESULT 6  
AAS46064  
ID AAS46064 standard; cDNA; 1636 BP.  
XX  
AC AAS46064;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Human DNA encoding PRO polypeptide sequence #140.  
XX  
KW PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep; ss;  
KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;  
KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;  
KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder;  
KW PCR primer.  
XX  
OS Homo sapiens.  
XX  
PN WO200168848-A2.  
XX  
PD 20-SEP-2001.  
XX  
PF 28-FEB-2001; 2001WO-US06520.  
XX  
PR 01-MAR-2000; 2000WO-US05601.  
PR 02-MAR-2000; 2000WO-US05841.  
PR 03-MAR-2000; 2000US-187202P.  
PR 06-MAR-2000; 2000US-186968P.  
PR 14-MAR-2000; 2000US-189320P.  
PR 14-MAR-2000; 2000US-189328P.  
PR 15-MAR-2000; 2000WO-US06884.  
PR 21-MAR-2000; 2000US-190828P.  
PR 21-MAR-2000; 2000US-191007P.  
PR 21-MAR-2000; 2000US-191048P.  
PR 21-MAR-2000; 2000US-191314P.  
PR 28-MAR-2000; 2000US-192655P.  
PR 29-MAR-2000; 2000US-193032P.  
PR 29-MAR-2000; 2000US-193053P.  
PR 30-MAR-2000; 2000WO-US08439.  
PR 04-APR-2000; 2000US-194449P.  
PR 11-APR-2000; 2000US-194647P.  
PR 11-APR-2000; 2000US-195975P.  
PR 11-APR-2000; 2000US-195975P.  
PR 11-APR-2000; 2000US-196000P.  
PR 11-APR-2000; 2000US-196187P.  
PR 11-APR-2000; 2000US-196690P.  
PR 11-APR-2000; 2000US-196820P.  
PR 18-APR-2000; 2000US-198121P.  
PR 18-APR-2000; 2000US-198585P.  
PR 25-APR-2000; 2000US-199397P.  
PR 25-APR-2000; 2000US-199550P.  
PR 25-APR-2000; 2000US-199654P.  
PR 03-MAY-2000; 2000US-201516P.  
PR 17-MAY-2000; 2000WO-US13705.  
PR 22-MAY-2000; 2000WO-US14042.  
PR 30-MAY-2000; 2000WO-US14941.  
PR 02-JUN-2000; 2000WO-US15264.  
PR 05-JUN-2000; 2000US-209832P.  
PR 28-JUL-2000; 2000WO-US20710.  
PR 22-AUG-2000; 2000US-0644848.  
PR 24-AUG-2000; 2000WO-US23328.  
PR 08-NOV-2000; 2000WO-US30952.  
PR 01-DEC-2000; 2000WO-US32678.  
PR 20-DEC-2000; 2000WO-US34956.  
XX

PA (GENTECH ) GENENTECH INC.

xx Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;  
PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;  
xx WPI: 2001-602746/68.

DR P-PSDB: AA029163.

xx Novel nucleic acids encoding PRO polypeptides, used to diagnose the  
PI presence of tumours, such as prostate and breast tumours, in mammals and  
PT to screen for modulators of the compounds -

PS Claim 2: Fig 279; 774pp: English.

xx Sequences AAS45925-AAS46231 represent DNA molecules encoding and PCR  
CC primers for PRO polypeptides of the invention. The sequences of the  
CC invention can be used to detect the presence of a tumour in a mammal by  
CC comparing the level of expression of a PRO polypeptide in a test sample  
CC of cells from the animal and a control sample of normal cells, whereby a  
CC higher level of expression in the test sample indicates the presence of a  
CC tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep,  
CC pigs, goats and rabbits but are preferably human. The polypeptides can be  
CC used to stimulate tumour necrosis factor (TNF) alpha release from human  
CC blood, when contacted with it. A specific polypeptide can be used to  
CC stimulate the proliferation or differentiation of chondrocyte cells. The  
CC PRO proteins can be used to determine the presence of tumours and also  
CC susceptibility to tumour development, particularly adrenal, lung, colon,  
CC breast, prostate, rectal, cervical, or liver tumours, in mammalian  
CC subjects. The oligonucleotide probes specific for the PRO nucleic acids  
CC can be used for genetic analysis of individuals with genetic disorders.

xx Sequence 1636 BP: 374 A: 492 C: 411 G: 359 T: 0 other:

Query Match 96.2%; Score 1617; DB 22; Length 1636;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 1628; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

QY 51 grggagcgcgccgagagcaccagctgcccaggtcgtgcatctctgtcctct 110  
DB 1 gaggagcgcgccgagagcaccagctgcccaggtcgtgcatctctgtcctct 60  
QY 111 gacacctgggaagatgcccgcgctggaccttaccctctctgtgttgcgtgcagc 170  
DB 61 gacacctgggaagatgcccgcgctggaccttaccctctctgtgttgcgtgcagc 120  
QY 171 caccctgtaccagcaccctcagttccacgtcagttcctcctcgcgccaaagtcac 230  
DB 121 caccctgtaccagcaccctcagttccacgtcagttcctcctcgcgccaaagtcac 180  
QY 231 caaagaagaagctgacacagaagctgaagacacaaagcaccagatctctgcagagct 290  
DB 181 caaagaagaagctgacacagaagctgaagacacaaagcaccagatctctgcagagct 240  
QY 291 gcccgtgtccacgtccatcgcggaagacccgagagc-atccctgtgctggcagcct 349  
DB 241 gcccgtgtccacgtccatcgcggaagacccgagagc-atccctgtgctggcagcct 300  
QY 350 ggtgaacaccgctcgtgaagacacatcctgtgctgaaggtcatalcaagctaaac 409  
DB 301 ggtgaacaccgctcgtgaagacacatcctgtgctgaaggtcatalcaagctaaac 360  
QY 410 gctgacagtggaagcctcgcgcaatgacagagctgctgacgtcagaatcccttgacat 469  
DB 361 gctgacagtggaagcctcgcgcaatgacagagctgctgacgtcagaatcccttgacat 420  
QY 470 ggtggtcgtgattcaaacacccctgtgcaagacatcgtggaattccacatgaagctga 529  
DB 421 ggtggtcgtgattcaaacacccctgtgcaagacatcgtggaattccacatgaagctga 480  
QY 530 ggcaccaagcaccatccgacatggaacacagtgcaagtgccccacccgctgtgtccacag 589  
DB 481 ggcaccaagcaccatccgacatggaacacagtgcaagtgccccacccgctgtgtccacag 540

QY 590 tgactgtgccaccagccatggagagcctgacatccaactgctgcataaagctctctcct 649  
DB 541 tgactgtgccaccagccatggagagcctgacatccaactgctgcataaagctctctcct 600  
QY 650 ggtgaacgctctagcacaagatgacagctcctcagtgccatccctcccaatagct 709  
DB 601 ggtgaacgctctagcacaagatgacagctcctcagtgccatccctcccaatagct 660  
QY 710 gaaaaacagctgtgtcccgatcgagagcttccctcaatgagatgacagacctct 769  
DB 661 gaaaaacagctgtgtcccgatcgagagcttccctcaatgagatgacagacctct 720  
QY 770 gacgctggtggaagtggtccatttccctcagcatgacgctcgtggaattgacctctgta 829  
DB 721 gacgctggtggaagtggtccatttccctcagcatgacgctcgtggaattgacctctgta 780  
QY 830 tctgtccatcaaggggtgacacatcagctcaccctggggccaaagtgtgtgactcaaca 889  
DB 781 tctgtccatcaaggggtgacacatcagctcaccctggggccaaagtgtgtgactcaaca 840  
QY 890 gggaaagtgaccacagtggttcaataactctgcagcttccctgacaaatgccaccctgga 949  
DB 841 gggaaagtgaccacagtggttcaataactctgcagcttccctgacaaatgccaccctgga 900  
QY 950 caacatccgcttcagctcatgtgagtgagagcgtggtgaaagctgacgtgtgtgt 1009  
DB 901 caacatccgcttcagctcatgtgagtgagagcgtggtgaaagctgacgtgtgtgt 960  
QY 1010 gctctctcagaagaatcagtgctcgtgtgacgtcgtgtcctcgtgaagtgccatcg 1069  
DB 961 gctctctcagaagaatcagtgctcgtgtgacgtcgtgtcctcgtgaagtgccatcg 1020  
QY 1070 gctgaagtcgaagcatcctggcgtgatcaatgaaaaagctgcagaaagctggtatccca 1129  
DB 1021 gctgaagtcgaagcatcctggcgtgatcaatgaaaaagctgcagaaagctggtatccca 1080  
QY 1130 gctcgtggaagatccttaaccacagacatcccgagttttatagaccaaagcgcatgccaa 1189  
DB 1081 gctcgtggaagatccttaaccacagacatcccgagttttatagaccaaagcgcatgccaa 1140  
QY 1190 ggtggtccacatgatatgctgtggaagtgttccctcctcagtgaaagcctccgacctgt 1249  
DB 1141 ggtggtccacatgatatgctgtggaagtgttccctcctcagtgaaagcctccgacctgt 1200  
QY 1250 caccctgggcacatcgaaagcagctcggaaagctcagtttcaaccaaaagtgaccacattat 1309  
DB 1201 caccctgggcacatcgaaagcagctcggaaagctcagtttcaaccaaaagtgaccacattat 1260  
QY 1310 actcaacttgaatcaatcagctcgtatgacacagctgatatgaactctgggaattggctg 1369  
DB 1261 actcaacttgaatcaatcagctcgtatgacacagctgatatgaactctgggaattggctg 1320  
QY 1370 gtcccaactgtatgttctgaaaaaacatcactcgtgacatcatccatccatccctgtgccc 1429  
DB 1321 gtcccaactgtatgttctgaaaaaacatcactcgtgacatcatccatccatccctgtgccc 1380  
QY 1430 gaacagaatgacaaatgaatcgtgggttccagtgatcattgtgtgaagccttggagatt 1489  
DB 1381 gaacagaatgacaaatgaatcgtgggttccagtgatcattgtgtgaagccttggagatt 1440  
QY 1490 cgaaggaagctgtgctcctcagacagagatgccccttggcttaccctcagcctcctgtg 1549  
DB 1441 cgaaggaagctgtgctcctcagacagagatgccccttggcttaccctcagcctcctgtg 1500  
QY 1550 gaaacccagctcctcgttccacagtgaaagactgtgatgacgacacacagcgaagcgtg 1609  
DB 1501 gaaacccagctcctcgttccacagtgaaagactgtgatgacgacacacagcgaagcgtg 1560  
QY 1610 gtcccaagctggagatggtgtgtgagctctatagaccatccctcctcgaatcaataaac 1669  
DB 1561 gtcccaagctggagatggtgtgtgagctctatagaccatccctcctcgaatcaataaac 1620



OY 1190 ggtggcccaactgacgtgctgtaagtgcttccctccagtgagccctccgcttctgt 1249  
|||||  
DB 1141 ggtggcccaactgacgtgctgtaagtgcttccctccagtgagccctccgcttctgt 1200  
OY 1250 caacctgggcatcgagagccagctcggaagctcagtttaacacaaagtgaccaattat 1309  
|||||  
DB 1201 caacctgggcatcgagagccagctcggaagctcagtttaacacaaagtgaccaattat 1260  
OY 1310 actcaactgtaatacatcagctgacgagtcgagtcagagtcggaactcctggatggctg 1369  
|||||  
DB 1261 actcaactgtaatacatcagctgacgagtcgagtcgagtcggaactcctggatggctg 1320  
OY 1370 gtccaacctgagtctgtcgtgaaacacatcactgagatcaccactccatctgtgccc 1429  
|||||  
DB 1321 gtccaacctgagtctgtcgtgaaacacatcactgagatcaccactccatctgtgccc 1380  
OY 1430 gaaccagaatgagcaaatlaagaatctcggggtcccaagtcgtaattggtggaagccttggatt 1489  
|||||  
DB 1381 gaaccagaatgagcaaatlaagaatctcggggtcccaagtcgtaattggtggaagccttggatt 1440  
OY 1490 cgaaggcaagtgcctcactcactgacaaagatgccttctgtctactcagcctctctgtg 1549  
|||||  
DB 1441 cgaaggcaagtgcctcactcactgacaaagatgccttctgtctactcagcctctctgtg 1500  
OY 1550 gaaacccagctctcctctgtctcccaagtgagacttgatgagcagcaccatcagggaagctgag 1609  
|||||  
DB 1501 gaaacccagctctcctctgtctcccaagtgagacttgatgagcagcaccatcagggaagctgag 1560  
OY 1610 gtccagactggagatgatgggtgtgagctctatagaccatccctctctgcaatcaataaac 1669  
|||||  
DB 1561 gtccagactggagatgatgggtgtgagctctatagaccatccctctctgcaatcaataaac 1620  
OY 1670 acttgccctgtg 1680  
|||||  
DB 1621 acttgccctgtg 1631

RESULT 8  
AAF54293  
ID AAF54293 standard; DNA; 1636 BP.  
XX  
AC AAF54293;  
XX  
DT 02-APR-2001 (first entry)  
XX  
DE DNA encoding protein of the invention #36.  
XX  
KW Secreted; transmembrane; gene therapy; ss.  
XX  
OS Unidentified.  
XX  
PN WO200078961-A1.  
XX  
PD 28-DEC-2000.  
XX  
PF 18-FEB-2000; 2000WO-US04342.  
XX  
PR 23-JUN-1999; 99US-0141037.  
PR 20-JUL-1999; 99US-0144758.  
PR 26-JUL-1999; 99US-0145698.  
PR 01-SEP-1999; 99WO-US20111.  
PR 29-OCT-1999; 99US-0162506.  
PR 30-NOV-1999; 99WO-US28313.  
PR 02-DEC-1999; 99WO-US28551.  
PR 16-DEC-1999; 99WO-US30095.  
PR 05-JAN-2000; 2000WO-US00219.  
PR 06-JAN-2000; 2000WO-US00376.  
XX  
PA (GENE ) GENENTECH INC.  
XX  
PI Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;  
PI Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;  
PI Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;

PI Watanabe CK, Williams PM, Wood WI;  
XX WPI: 2001-071395/08.  
DR  
XX Secreted and transmembrane proteins and nucleic acids designated PRO,  
PT useful as hybridization probes, in chromosome and gene mapping and gene  
PT therapy -  
XX  
XX  
PS Claim 2; Fig 71; 787pp; English.  
XX  
XX The present invention relates to secreted and transmembrane proteins.  
CC These proteins and the DNA encoding them may be used as hybridization  
CC probes, in chromosome and gene mapping and in the generation of  
CC anti-sense RNA and DNA. They may also be used to generate either  
CC transgenic animals or knockout animals which are in turn useful for  
CC development and screening of therapeutically useful reagents.  
XX The nucleic acids may also be used in gene therapy.  
XX  
SQ Sequence 1636 BP; 374 A; 492 C; 411 G; 359 T; 0 other;

Query Match 96.2%; Score 1617; DB 22; Length 1636;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1628; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

OY 51 grggagcgggcgccgagagactccagcgtgcccaggtctgagatctgacttgcctctct 110  
|:|||||  
DB 1 gagggagcgggcgccgagagactccagcgtgcccaggtctgagatctgacttgcctctct 60  
OY 111 gacacctgggaagatgagccgcccgtgagacttcaacctctctctgtgttctgagcagc 170  
|||||  
DB 61 gacacctgggaagatgagccgcccgtgagacttcaacctctctctgtgttctgagcagc 120  
OY 171 caacttgatccaagccacctcagttcccaactgcaattcattcctccggcccaaaagtcac 230  
|||||  
DB 121 caacttgatccaagccacctcagttcccaactgcaattcattcctccggcccaaaagtcac 180  
OY 231 caaagaagaagctgacacaggaagctgaagaccacacagccacagatcctcagcagct 290  
|||||  
DB 181 caaagaagaagctgacacaggaagctgaagaccacacagccacagatcctcagcagct 240  
OY 291 gccgctgctcagtgctcagtcgggaaagccagccggagag-alcctctgtctgaggaccc 349  
|||||  
DB 241 gccgctgctcagtgctcagtcgggaaagccagccggagagatcctctgtctgaggaccc 300  
OY 350 ggtgaacacgctccctggaagacatcatctgctggaagtcacatcagcttaaatctcca 409  
|||||  
DB 301 ggtgaacacgctccctggaagacatcatctgctggaagtcacatcagcttaaatctcca 360  
OY 410 gctcaggtgaaagccctcgcccaatgacagcagctgctcagtcgaatctcccttgagac 469  
|||||  
DB 361 gctcaggtgaaagccctcgcccaatgacagcagctgctcagtcgaatctcccttgagac 420  
OY 470 ggtgctgagatccaacacgcccctgtgtcaagaccatcgtgaggttccacatgacactga 529  
|||||  
DB 421 ggtgctgagatccaacacgcccctgtgtcaagaccatcgtgaggttccacatgacactga 480  
OY 530 ggcccaagccacatccgcatggaacacagtgcaagtggtgcccacccgctgtgctccag 589  
|||||  
DB 481 ggcccaagccacatccgcatggaacacagtgcaagtggtgcccacccgctgtgctccag 540  
OY 590 tgactgtgcccacacgacatggagagctgtgacatccaacgtcgcatagctctctctct 649  
|||||  
DB 541 tgactgtgcccacacgacatggagagctgtgacatccaacgtcgcatagctctctctct 600  
OY 650 ggtgaagccttagtctaagcaggtcatggaacctctagtgtccatccctcccaatcagt 709  
|||||  
DB 601 ggtgaagccttagtctaagcaggtcatggaacctctagtgtccatccctcccaatcagt 660  
OY 710 gaaacacagcgtgtgtcccgtagatcgaagcttccctcaatggcatgtatgagacctct 769  
|||||  
DB 661 gaaacacagcgtgtgtgtcccgtagatcgaagcttccctcaatggcatgtatgagacctct 720



Db 716 gagctgaaggaccacaagccacagcatcctcgcagcagctgcgcgtctcattgcatg 775  
Oy 310 cgggaaaaagccagccggaggg -atccctgtgctggcagcctgtgtaaacacgctctgaag 368  
Db 776 cgggaaaaagccagccggagggcatccctgtgctggcagcctgtgtaaacacgctctgaag 835  
Oy 369 cacatcatctgtgctgaaggtcatcatcaagctaaacatccctcagctgcaggtgaagccctcg 428  
Db 836 cacatcatctgtgctgaaggtcatcatcaagctaaacatccctcagctgcaggtgaagccctcg 895  
Oy 429 gccaatgacccaggaagctcgtatgaagaatccctcgtgaacatggtgctgtgatcaacag 488  
Db 896 gccaatgacccaggaagctcgtatgaagaatccctcgtgaacatggtgctgtgatcaacag 955  
Oy 489 cccctggtgcaagaacatcgtgaggtgtccatagacgactggaagcccaagccacatccgc 548  
Db 956 cccctggtgcaagaacatcgtgaggtgtccatagacgactggaagcccaagccacatccgc 1015  
Oy 549 atggaacccagtgcaaggtgccccaccccgctgctccatgtaactgtgccccacagccat 608  
Db 1016 atggaacccagtgcaaggtgccccaccccgctgctccatgtaactgtgccccacagccat 1075  
Oy 609 ggggaagctgcgcataccaactcgtcgcataagctctccttcgtggtgaagccttagctaa 668  
Db 1076 ggggaagctgcgcataccaactcgtcgcataagctctccttcgtggtgaagccttagctaa 1135  
Oy 669 caggtcatgaacctctctagctccatccctgcgcatactagtgaaacacagctgtgtccc 728  
Db 1136 caggtcatgaacctctctagctccatccctgcgcatactagtgaaacacagctgtgtccc 1195  
Oy 729 gtgatcggagcttccctcaatgcatgtatgcagacccctcctgcacatggtgtgaaggtgcc 788  
Db 1196 gtgatcggagcttccctcaatgcatgtatgcagacccctcctgcacatggtgtgaaggtgcc 1255  
Oy 789 atttccctcagcatggaacgctcgtgagtttgacctctgtaacctgcatacaagggctaac 848  
Db 1256 atttccctcagcatggaacgctcgtgagtttgacctctgtaacctgcatacaagggctaac 1315  
Oy 849 accatcagcctctacctcgtggggcgaagttgttggaactcaacagggaaggtgacgaagtg 908  
Db 1316 accatcagcctctacctcgtggggcgaagttgttggaactcaacagggaaggtgacgaagtg 1375  
Oy 909 ttcaataactctgcagcttccctgcgaatgccccacctggaacaaatcccgcttcagcctc 968  
Db 1376 ttcaataactctgcagcttccctgcgaatgccccacctggaacaaatcccgcttcagcctc 1435  
Oy 969 atcgtgagtcaggaagcgtgtgaagaactgcagtgctgctgctctctccagaagaattc 1028  
Db 1436 atcgtgagtcaggaagcgtgtgaagaactgcagtgctgctgctctctccagaagaattc 1495  
Oy 1029 atggtccgtgtggaactcgtctctcgtgagagtgcccaatcggtgaagtcaagcatcg 1088  
Db 1496 atggtccgtgtggaactcgtctctcgtgagagtgcccaatcggtgaagtcaagcatcg 1555  
Oy 1089 ctgatcaatgaaaaagcgtcgcagataagctggaatctaacccaatcgttaagaatcctaact 1148  
Db 1556 ctgatcaatgaaaaagcgtcgcagataagctggaatctaacccaatcgttaagaatcctaact 1615  
Oy 1149 caggaacactcccgag -tttttatagaccaagggcatalgccaaggtgccccaaactgatacgt 1207  
Db 1616 caggaacactcccgagtttttatagaccaagggcatalgccaaggtgccccaaactgatacgt 1675  
Oy 1208 gctggaagtggttccctcagtggaagccctcgcgccttggttcaaccctggggatgtgaagc 1267  
Db 1676 gctggaagtggttccctcagtggaagccctcgcgccttggttcaaccctggggatgtgaagc 1735  
Oy 1268 cagctcggaaagctcagttttaaaccaaaaggtgacaaacttataactcaactgtgaataaac 1327  
Db 1736 cagctcggaaagctcagttttaaaccaaaaggtgacaaacttataactcaactgtgaataaac 1795  
Oy 1328 cagctcgtatcggatccagctgtgaactcgtggaattggtggttccaaactgtatgttct 1387  
Db 1796 cagctcgtatcggatccagctgtgaactcgtggaattggtggttccaaactgtatgttct 1855

Oy 1388 gaaaaacatcatcatgagatcatcatccactccactcgtcgtccgaacagaatgccaatt 1447  
Db 1856 gtaaaacatcatcatgagatcatcatccactccactcgtcgtccgaacagaatgccaatt 1915  
Oy 1448 aagatctgggtgtccagatgctatggtgaagagccttgggaltcgaagcagctgaatcctc 1507  
Db 1916 aagatctgggtgtccagatgctatggtgaagagccttgggaltcgaagcagctgaatcctc 1975  
Oy 1508 actgaccaaagatgctcctgtgtcttaactccagccctcctgtggaaccacagctccctgt 1567  
Db 1976 actgaccaaagatgctcctgtgtcttaactccagccctcctgtggaaccacagctccctgt 2035  
Oy 1568 ctcccaagtgaagacttgatgagcagcatcaagggaagcgtgggtccagctggagatg 1627  
Db 2036 ctcccaagtgaagacttgatgagcagcatcaagggaagcgtgggtccatctcgtgagatg 2095  
Oy 1628 ggtgtgagctctatagaccatccct 1652  
Db 2096 ggtgtgagctctatagaccatccct 2120

RESULT 10  
ABA01989  
ID ABA01989 standard; cDNA: 1035 BP.  
XX  
AC ABA01989;  
XX  
DT 12-FEB-2002 (first entry)  
XX  
DE Human NOV6b coding sequence.  
XX  
KW Human; NOV6b; metabolic disorder; neurodegenerative disorder;  
KW immune disorder; haematopoietic disorder; developmental disease; cancer;  
KW retinal disease; feeding disorder; vaccine; infection; gene therapy;  
KW neurological disorder; psychotic disorder; G-protein coupled receptor;  
KW cytoskeletal; antidiabetic; virucide; neuroprotective; nootropic;  
KW analgesic; antidepressant; antimigraine; anticonvulsant; neuroleptic;  
KW antihistaminic; antiallergic; antiinflammatory; anorectic; antiarthritic;  
KW antipsoriatic; antithrombotic; antibacterial; fungicide;  
KW osteopathic; protozoacide; antiulcer; hypertensive; hypotensive;  
KW antiinfertility; vulnerrary; nephrotoxic; antileptic;  
KW salivary gland protein; chromosome 20; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
FH Key Location/Qualifiers  
FT 1..79  
FT 5'UTR  
FT /\*tag= a  
FT 79..1035  
FT CDS  
FT /\*tag= b  
FT /product= "NOV6b"  
FT sig\_peptide  
FT 79..150  
FT /\*tag= c  
FT mat\_peptide 151..1032  
FT /\*tag= d  
FN WO200174851-A2.  
PN  
XX  
PD 11-OCT-2001.  
XX  
XX  
PF 30-MAR-2001; 2001WO-US10039.  
PR 30-MAR-2000; 2000US-193205P.  
PR 30-MAR-2000; 2000US-193339P.  
PR 05-APR-2000; 2000US-195343P.  
PR 06-APR-2000; 2000US-195005P.  
PR 06-APR-2000; 2000US-195088P.  
PR 10-APR-2000; 2000US-195792P.  
PR 11-APR-2000; 2000US-196556P.  
PR 13-APR-2000; 2000US-197081P.  
PR 14-APR-2000; 2000US-197087P.  
PR 14-APR-2000; 2000US-197525P.

PR 30-MAR-2000; 2000US-193205P  
PR 30-MAR-2000; 2000US-193339P  
PR 05-APR-2000; 2000US-195343P  
PR 06-APR-2000; 2000US-195005P



Query Match	43.18;	Score 724.2;	DB 22;	Length 963;
Best Local Similarity	90.18;	Pred. No. 2.7e-178;		
Matches 842; Conservative	0;	Mismatches 38;	Indels 55;	Gaps 4;

Db	481	gcccacacgcccattgagagccttcgcgcatccaaactgctgataagcctctcttcgttgaac	540
Oy	657	gccttagagtaaggaagatataaactctcttaagtacatccctcgcccaatctagtgaanaac	716
Db	541	gccttaagctaaagcaagatataaactctcttaagtacatccctcgcccaatctagtgaanaac	600
Oy	717	cagctgtctcccgatcatagagccttccctcaatgagatataagacactctctcagctg	776
Db	601	cagctgtctcccgatcatagagccttccctcaatgagatataagacactctctcagctg	660
Oy	777	gtgaaggtgcccacattccctcagcatggaaccgtctggaatttgaaactctgtatccctgcc	836
Db	661	gtgaaggtgcccacattccctcagcatggaaccgtctggaatttgaaactctgtatccctgcc	700
Oy	837	atcaagaggtgacacacattcagctcatctactctggtggggcgaagttgtgtgactcaaggaag	896
Db	701	-----	732
Oy	897	gtgaccagatgtgtcaataactctgacagcttccctcctgacaatgcccacccctggaacatc	956
Db	733	gtgaccagatgtgtgtcaataactctgacagcttccctcctgacaatgcccacccctggaacatc	792
Oy	957	ccgttcacagcctcatcgttgatgcaggaagctgtgtgaagctgcagctgctgtctctct	1016
Db	793	ccgttcacagcctcatcgttgatgcaggaagctgtgtgaagctgcagctgctgtctctct	852
Oy	1017	ccagaagaattcatcagtgctcctgttggactctgtgct	1051
Db	853	ccagaagaattcatcagtgctcctgttggactctgtgct	887
RESULT_12			
AA297114			
ID	AA297114	standard; cDNA; 1172 BP.	
XX	AA297114:		
XX	19-APR-2000	(first entry)	
DE	Human secreted protein gene 5 cDNA clone HBHMA23, SEQ ID NO:106.		
KW	Human; secreted protein; cancer; tumour; developmental abnormality;		
KW	foetal deficiency; blood disorder; immune system disorder; inflammation;		
KW	autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;		
KW	schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;		
KW	atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;		
KW	digestive disorder; endocrine disorder; infection; AIDS; leukaemia;		
XX	therapy; ds.		
XX	OS	Homo sapiens.	
XX	PN	W09966041.A1.	
XX	PD	23-DEC-1999.	
XX	PE	15-JUN-1999; 99MO-US13418.	
XX	PR	16-JUN-1998; 98US-0089507.	
XX	PR	16-JUN-1998; 98US-0089508.	
XX	PR	16-JUN-1998; 98US-0089509.	
XX	PR	16-JUN-1998; 98US-0089510.	
XX	PR	22-JUN-1998; 98US-0090112.	
XX	PR	22-JUN-1998; 98US-0090113.	
XX	PA	(HUMA-) HUMAN GENOME SCI INC.	
XX	PI	Ruben SM, Ni J, Rosen CA, Wei Y, Young PE, Florence KA;	
XX	PI	Soppet DR, Brewer LA, Endress GA, Carter KC, Muenesi M, Ebner R;	
XX	PI	Lafleur DM, Olsen HS, Shi Y, Moore PA, Komatsoulis G;	
XX	PI	WPI; 2000-106100/09.	
DR	P-PSDB; AAY86310.		

XX New isolated human genes and the secreted polypeptides they encode,  
PT useful for diagnosis and treatment of e.g. cancers, neurological  
PT disorders, immune diseases, inflammation or blood disorders -  
XX  
XX  
PS Claim 1: Page 374; 586pp; English.

CC AA97019 to AA97137 represent 94 isolated human secreted protein genes.  
CC AA96215 to AA96333 are the secreted proteins encoded by the 94 human  
CC genes. This sequence represents a fragment of one of the human secreted  
CC proteins. The genes and their corresponding secreted polypeptides are  
CC useful for preventing, treating or ameliorating medical conditions,  
CC e.g., by protein or gene therapy. Also pathological conditions can be  
CC diagnosed by determining the presence of mutations in the new genes. Specific  
CC uses are described for each of the 94 genes, based on which tissues they  
CC are most highly expressed in, and include developing products for the  
CC diagnosis or treatment of cancer, tumours, developmental abnormalities  
CC and foetal deficiencies, blood disorders, diseases of the immune system,  
CC autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive  
CC disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin  
CC disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney  
CC disorders, digestive/endoctrine disorders, infections and AIDS. The  
CC polypeptides are also useful for identifying their binding partners.  
CC The sequences shown in AA96334 to AA96585 represent fragments of the  
CC secreted proteins.  
CC  
XX  
SQ Sequence 1172 BP; 287 A; 358 C; 281 G; 238 T; 8 other;

Query Match 38.2%; Score 641; DB 21; Length 1172;  
Best Local Similarity 99.1%; Pred. NO. 1.3e-156;  
Matches 641; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 56 gcggcgagagactccagcgtgcccaggtctgcatctcgtacgtgaccttgacac 115  
DB 2 gcggcgagagactccagcgtgcccaggtctgcatctcgtacgtgaccttgacac 61  
QY 116 ctgggaagatggtccgcccgtggaactcaccctctctgtgttctggtgacgacact 175  
DB 62 ctgggaagatggtccgcccgtggaactcaccctctctgtgttctggtgacgacact 121  
QY 176 tgatccaaagcaccctcagtcaccctcagttcattcattcctcgtgcccagaatcaag 235  
DB 122 tgatccaaagcaccctcagtcaccctcagttcattcattcctcgtgcccagaatcaag 181  
QY 236 aaaaagtacacagagactgaagacacacacagccacagatcctcgaagctgcgc 295  
DB 182 aaaaagtacacagagactgaagacacacacagccacagatcctcgaagctgcgc 241  
QY 296 tgcctaatgcatcgtgggaaagcagccgagagatccctgtgtctggtgacgacctgtgaa 355  
DB 242 tgcctaatgcatcgtgggaaagcagccgagagatccctgtgtctggtgacgacctgtgaa 301  
QY 356 caccgtcctaaagacacatcttgcgtgaaggtcatcacagctcctaactcctcagctgca 415  
DB 302 caccgtcctaaagacacatcttgcgtgaaggtcatcacagctcctaactcctcagctgca 361  
QY 416 ggtgaagcctcgtgcaatgaacagagagctgtactgaatcaccctctggaacatgtgac 475  
DB 362 ggtgaagcctcgtgcaatgaacagagagctgtactgaatcaccctctggaacatgtgac 421  
QY 476 tggattcaaacagccctctgtaagaacacatcgtgagattcacaatagcactagagccca 535  
DB 422 tggattcaaacagccctctgtaagaacacatcgtgagattcacaatagcactagagccca 481  
QY 536 agccacacatccgcatggaacccagctgaagtggcccaaccgctgtcctcagctgac 595  
DB 482 agccacacatccgcatggaacccagctgaagtggcccaaccgctgtcctcagctgac 541  
QY 596 tgcacacagcactggtgagcctcgtcgaatcacaatcgtgcatataagctcctcctgtgtaa 655  
DB 542 tgcacacagcactggtgagcctcgtcgaatcacaatcgtgcatataagctcctcctgtgtaa 601

QY 656 cgccttagtaagcaggtcatgaactcctcgtgcatcctcctgcca 702  
DB 602 cgccttagtaagcaggtcatgaactcctcgtgcatcctcctgcca 648

RESULT 13  
AAFI8254  
ID AAF18254 standard; DNA; 1177 BP.  
AC AAF18254;  
XX 14-MAR-2001 (first entry)  
XX  
XX Lung cancer associated polynucleotide sequence SEQ ID 273.

KW Human; lung cancer associated protein; neuroprotective; cytoskeletal;  
KW cardioactive; immunomodulatory; muscular active; vulneryary;  
KW gastrointestinal; nephrotropic; antiinfective; gynecological;  
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;  
KW proliferative disorder; wound healing; infectious disease; ds.

OS Homo sapiens.  
XX  
XX WO20005180-A2.  
XX  
XX 21-SEP-2000.  
XX  
XX 08-MAR-2000; 2000MO-US05918.  
XX  
XX 12-MAR-1999; 9905-0124270.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX (ROSE/) ROSEN C A.

XX Ruben SM;  
XX  
XX WPI; 2000-587514/55.  
XX  
XX P-P-SDB; AAB58378.

PT Lung cancer associated gene sequences, referred to as lung cancer  
PT actigens, useful for treatment, prevention, and diagnosis of disorders  
PT such as lung cancer -  
XX  
XX  
XX Claim 1: Page 732; 1425pp; English.

CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer  
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer  
CC associated proteins and polynucleotide sequences, their agonists, and  
CC antagonists may have neuroprotective; cytoskeletal; cardioactive;  
CC immunomodulatory; muscular active general; vulneryary; gastrointestinal  
CC general; nephrotropic; antiinfective; gynecological; or antibacterial  
CC activity. The invention also includes antibodies specific for the  
CC protein or polynucleotide sequences. The lung cancer associated  
CC polynucleotide sequences may be used for detection of lung cancer,  
CC chromosome identification, as chromosome markers, and for numerous other  
CC diagnostic or research purposes. The proteins may be used to treat  
CC disorders such as neural, immune, muscular, reproductive,  
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative  
CC disorders. The proteins may also be used in the treatment of wounds and  
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and  
CC peptide AAB58549 are used in the course of the invention for the  
CC identification and characterisation of the polynucleotide and protein  
CC sequences.  
XX  
XX

Sequence 1177 BP; 290 A; 361 C; 287 G; 239 T; 0 other;

Query Match 37.8%; Score 635.8; DB 21; Length 1177;  
Best Local Similarity 99.5%; Pred. NO. 2.9e-155;  
Matches 648; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 53 ggaagcgagcagagactccagcgtgcccaggtctgcatcctcgtacgtgacctgtgcctctga 112

[illegible]

	PR	16-JUN-1998;	98US-0089510.	
	PR	22-JUN-1998;	98US-0090112.	
	PR	22-JUN-1998;	98US-0090113.	
	PX	(HUMA - ) HUMAN GENOME SCI INC.		
	PA	Ruben SM, Ni J, Rosen CA, Mei Y, Young PE, Florence KA; Pi Soppet DR, Brewer LA, Endress GA, Carter KC, Nicenski M, Pl Latleur DW, Olsen HS, Shi Y, Moore PA, Komatsoulis G;	Eder R;	
	PX	WP1; 2000-106100/09. P-PDB: AAY86219.		
	DR	New isolated human genes and the secreted polypeptides they encode,		
	PX	useful for diagnosis and treatment of e.g. cancers, neurological disorders, immune diseases, inflammation or blood disorders -		
	PT	Claim 1; Page 318; 586pp; English.		
	PX	AAS97019 to AAZ97137 represent 94 isolated human secreted protein genes.		
	CC	AAY86215 to AAY86333 are the secreted proteins encoded by the 94 human genes. This sequence represents a fragment of one of the human secreted proteins. The genes and their corresponding secreted polypeptides are useful for preventing, treating or ameliorating medical conditions, e.g., by protein or gene therapy. Also pathological conditions can be diagnosed by determining the presence of mutations in the new genes. Specific uses are described for each of the 94 genes, based on which tissues they are most highly expressed in, and include developing products for the diagnosis or treatment of cancer, tumours, developmental abnormalities and foetal deficiencies, blood disorders, diseases of the immune system, autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin disorders, scleroderma, diabetes, cardiovascular disorders, kidney disorders, digestive/endocrine disorders, infections and AIDS. The polypeptides are also useful for identifying their binding partners. The sequences shown in AAY86334 to AAY86585 represent fragments of the secreted proteins.		
	CC	SQ Sequence 1175 BP; 290 A; 359 C; 286 G; 239 T; 1 other:		
	XX			
	Query Match	37.8%; Score 634.8; DB 21; Length 1175;		
	Best Local Similarity	99.5%; Pred. No. 5,2e+155;		
	Matches 647; Conservative	0; Mismatches 2; Indels 1; Gaps 1		
OY	54	gagcgggccgaaggactcagaagtgcgccaaagtcttgatattctgtacctgtgccccctttagac	113	
Dd	1	gagcgggccgaaggactcagaagtgcgccaaagtcttgatattctgtacctgtgccccctttagac	60	
OY	114	accctgggaagatgagcccggccgttgacaaccttaaccccttccttgtgtttctgtgcaagcac	173	
Dd	61	accctgggaagatgagcccggccgttgacaaccttaaccccttccttgtgtttctgtgcaagcac	120	
OY	174	cctgatccaagccaccctcagtcaccaactgcatlctcatactcctctggccaaaagtatacaa	233	
Dd	121	cctgatccaagccaccctcagtcaccaactgcatlctcatactcctctggccaaaagtatacaa	180	
OY	234	agaaaaagctgacaacagagatctaagagaccacaacagccaccacagatatcttcagcagattgac	293	
Dd	181	agaaaaagctgacaacagagatctaagagaccacaacagccaccacagatatcttcagcagattgac	240	
OY	294	gctgtctcattgcatatgcgggaaaaaacgcagccggaagt-atccctgtgtctggcagccttgtt	352	
Dd	241	gctgtctcattgcatatgcgggaaaaaacgcagccggaagt-atccctgtgtctggcagccttgtt	300	
OY	353	gaacacccgtctcgaagacaatatcatctgtgtttaaggltatalcacagtaaataatccctcagat	412	
Dd	301	gaacacccgtctcgaagacaatatcatctgtgtttaaggltatalcacagtaaataatccctcagat	360	
OY	413	gcaggttgaaagccctcggccaatgataccagagatctgtagtaagaattcccttcygaacatggt	472	
Dd	361	gcaggttgaaagccctcggccaatgataccagagatctgtagtaagaattcccttcygaacatggt	420	

QY 473 ggcctgattcaacacgcccctggtcaagaccatcgtgagttccacatgaagcagc 532  
 Db 421 ggcctgattcaacacgcccctggtcaagaccatcgtgagttccacatgaagcagc 480  
 QY 533 ccaagccaccatccgcatgagacacagtcgaagtcgcccacccgctggtctcaatga 592  
 Db 481 ccaagccaccatccgcatgagacacagtcgaagtcgcccacccgctggtctcaatga 540  
 QY 553 ctgtgcccacacagcattgagcctgcatcccaactgctgcatgaagctctctctcgt 652  
 Db 541 ctgtgcccacacagcattgagcctgcatcccaactgctgcatgaagctctctctcgt 600  
 QY 653 gaagccttaagcagcagtcatacctcctcctagtcacccctcctcctcctcctc 702  
 Db 601 gaagccttaagcagcagtcatacctcctcctagtcacccctcctcctcctcctc 650

RESULT 15

AAH30554 standard; cDNA; 420 BP.

AAH30554;

27-JUL-2001 (first entry)

Human colon cancer cell line Km12L4-A cDNA library derived sequence #488.

Human; diagnosis: colon cancer; cancer: malignant; chromosome mapping; detection: colon cancer cell line Km12L4-A, ss.

Homo sapiens.

WO200018916-A2.

06-APR-2000.

23-SEP-1999; 99WO-US22226.

28-SEP-1998; 98US-0102161.

28-SEP-1998; 98US-0102180.

29-SEP-1998; 98US-0102380.

08-OCM-1998; 98US-0103815.

27-OCM-1998; 98US-0105877.

(CHIR) CHIRON CORP.

(HYSE-) HYSEO INC.

Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;

Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;

Iamson G, Drmanac R, Ckvenjakov R, Dickson M, Drmanac S, Labat I;

Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;

WPI; 2000-293155/25.

Polynucleotide library comprising 1079 defined sequences, useful in

the form of an array to detect cancer or susceptibility to cancer -

Claim 1; Page 323-324; 502pp; English.

The present invention describes a library of polynucleotides comprising 1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described are: (1) an isolated polynucleotide (I) having at least 90% identity to one of the 1079 sequences; (2) a recombinant host cell containing (I); (3) an isolated polypeptide (II) encoded by (I); (4) an antibody that specifically binds to (II); (5) a vector comprising (I); and (6) a method of detecting differentially expressed genes correlated with a cancerous state of a mammalian cell comprising detecting a gene product encoded by 65 of the 1079 sequences given in the specification. The polynucleotides are used to monitor patients having (or susceptible) to cancer to detect potentially malignant events at a molecular level before they are detectable at a gross morphological level. The polynucleotides are also useful for monitoring the efficacy of various therapies and preventive

CC interventions. Polynucleotide probes based on the disclosed sequences  
 CC are useful for chromosome mapping and detection of transcription levels.  
 CC The 1079 polynucleotide sequences were derived from a human colon cancer  
 CC cell line Km12L4-A cDNA library.

XX Sequence 420 BP; 97 A; 116 C; 108 G; 99 T; 0 other;

Query Match 25.0%; Score 420; DB 21; Length 420;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-99;  
 Matches 420; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 870 gccaatgttgagctcaacaggaagtgaccagtggttcaatgaactcgcagctcc 929  
 Db 1 gccaatgttgagctcaacaggaagtgaccagtggttcaatgaactcgcagctcc 60  
 QY 930 ctgacaatgcccacccctggaacaacatccglttcagctcactgtagtcagagcgtgtg 989  
 Db 61 ctgacaatgcccacccctggaacaacatccglttcagctcactgtagtcagagcgtgtg 120  
 QY 990 aaagctgagtgctgt 1049  
 Db 121 aaagctgagtgctgt 180  
 QY 1050 ctctcctgagagtgccatcgctgaagtgacagcagtcggtgataatgaagaagctgca 1109  
 Db 181 ctctcctgagagtgccatcgctgaagtgacagcagtcggtgataatgaagaagctgca 240  
 QY 1110 gataagctggatctaccagatcgatgaagatccttaactcagagacatcccgagtttt 1169  
 Db 241 gataagctggatctaccagatcgatgaagatccttaactcagagacatcccgagtttt 300  
 QY 1170 atagaccagagcagtcgcaagtggtgcccacatgatgtgtgtgaagtgttccctcagt 1229  
 Db 301 atagaccagagcagtcgcaagtggtgcccacatgatgtgtgtgaagtgttccctcagt 360  
 QY 1230 gaagccctccgccccttggttcacccctgggcatggaagccagcgcggaagctcagtttac 1289  
 Db 361 gaagccctccgccccttggttcacccctgggcatggaagccagcgcggaagctcagtttac 420

Search completed: September 16, 2002, 02:55:06  
 Job time: 7384 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 00:51:17 : Search time 89 Seconds  
(without alignments)  
4636.677 Million cell updates/sec

Title: US-09-700-770-2  
Perfect score: 1680  
Sequence: 1 ggtgtgagagatataagtt.....tcaataaacacttgctgtg 1680

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_NA:  
1: /cgn2\_6/ptodata/2/1na/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/1na/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/1na/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/1na/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/1na/PCUTS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/1na/Backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41.6	2.5	44377	2	US-08-804-227C-7
2	41.6	2.5	44377	2	US-08-804-198-1
3	39.6	2.4	4245	4	US-09-276-531-16
C	38.4	2.3	4403765	4	US-09-103-840A-2
5	38.2	2.3	1931	2	US-09-130-114-2
6	37.6	2.2	1301	4	US-08-804-166-7
7	37.6	2.2	1301	4	US-08-910-991-7
8	37.4	2.2	3241	4	US-09-434-288-11
9	37.2	2.2	1241	1	US-08-471-033-39
10	37.2	2.2	1241	1	US-08-471-033-42
11	37.2	2.2	1241	2	US-08-471-044-39
12	37.2	2.2	1241	2	US-08-471-044-42
13	37.2	2.2	1241	2	US-08-463-483A-39
14	37.2	2.2	1241	2	US-08-463-483A-42
15	37.2	2.2	1241	2	US-08-471-046A-39
16	37.2	2.2	1241	2	US-08-471-046A-42
17	37.2	2.2	1241	2	US-08-470-566B-39
18	37.2	2.2	1241	2	US-08-470-566B-42
19	37.2	2.2	1241	2	US-08-469-334-39
20	37.2	2.2	1241	2	US-08-469-334-42
21	37.2	2.2	1241	3	US-09-300-529-39
22	37.2	2.2	1241	3	US-09-300-529-42
23	37.2	2.2	1358	1	US-08-471-033-45
24	37.2	2.2	1358	2	US-08-471-044-45
25	37.2	2.2	1358	2	US-08-463-483A-45
26	37.2	2.2	1358	2	US-08-471-046A-45
27	37.2	2.2	1358	2	US-08-470-566B-45

28	37.2	2.2	1358	2	US-08-469-334-45	Sequence 45, Appl
29	37.2	2.2	1358	3	US-09-300-529-45	Sequence 45, Appl
30	37.2	2.2	1389	1	US-08-471-033-27	Sequence 27, Appl
31	37.2	2.2	1389	2	US-08-471-044-27	Sequence 27, Appl
32	37.2	2.2	1389	2	US-08-463-483A-27	Sequence 27, Appl
33	37.2	2.2	1389	2	US-08-471-046A-27	Sequence 27, Appl
34	37.2	2.2	1389	2	US-08-470-566B-27	Sequence 27, Appl
35	37.2	2.2	1389	2	US-08-469-334-27	Sequence 27, Appl
36	37.2	2.2	1389	3	US-09-300-529-27	Sequence 27, Appl
C	37.2	2.2	1995	1	US-08-425-063-3	Sequence 3, Appl
38	37.2	2.2	1995	2	US-08-317-844B-3	Sequence 3, Appl
39	37.2	2.2	4031	1	US-08-471-033-49	Sequence 49, Appl
40	37.2	2.2	4031	2	US-08-471-044-49	Sequence 49, Appl
41	37.2	2.2	4031	2	US-08-463-483A-49	Sequence 49, Appl
42	37.2	2.2	4031	2	US-08-471-046A-49	Sequence 49, Appl
43	37.2	2.2	4031	2	US-08-470-566B-49	Sequence 49, Appl
44	37.2	2.2	4031	2	US-08-469-334-49	Sequence 49, Appl
45	37.2	2.2	4031	3	US-09-300-529-49	Sequence 49, Appl

ALIGNMENTS

RESULT 1  
US-08-804-227C-7  
; Sequence 7, Application US/08804227C  
; Patent No. 5876991  
; GENERAL INFORMATION:  
; APPLICANT: Dehoff, Bradley S.  
; APPLICANT: Kuhstoss, Stuart A.  
; APPLICANT: Rostock, Paul R., Jr.  
; APPLICANT: Sutton, Kimberly L.  
; TITLE OF INVENTION: POLYPEPTIDE SYNTHASE GENES  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: THOMAS G. PLANT 1501  
; STREET: LILLY CORPORATE CENTER  
; CITY: INDIANAPOLIS  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: ASCII(DOS) Text, only  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/804,227C  
; FILING DATE: February 21, 1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plant, Thomas, G.  
; REGISTRATION NUMBER: 35,784  
; REFERENCE/DOCKET NUMBER: X-8231  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-2459  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 44377 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 350..14002  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 14046..20036  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 20110..31284  
; FEATURE:

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.22  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/276,531  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/079,677  
FILING DATE: March 27, 1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Lynn E. Murry, Ph.D.

```
;
;
;   REGISTRATION NUMBER: 42,918
;   REFERENCE/DOCKET NUMBER: PA-0008 US
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (650) 855-0555
;   TELEFAX: (650) 845-4166
;   INFORMATION FOR SEQ ID NO: 16:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 4245 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   IMMEDIATE SOURCE:
;     LIBRARY: BRAITUT21
;     CLONE: 2526432
;   US-09-276-531-16

Query Match
Best Local Similarity 43.8%; Score 39.6; DB 4; Length 4245;
Matches 171; Conservative 0; Mismatches 219; Indels 0; Gaps 0;

QY 357 accgtcttaagacatcatctctgctggaaggtcattacagctaaactcttcagctgcag 416
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 130 AGCTTCACATCGGACATCGGCCCGTACACAGCGGGCGCTTCTACACCAAGCCACCTCAG 189

QY 417 gtgaagccttcggaacacagagagctgtagtaagatcccccctgtaacatggtggtc 476
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 190 TGGGTGGACATCCCGCGGACCTCGCGCTGTGCGCACACGTGGCTACAGAGAGATGCTG 249

QY 477 ggaattacaagcgccttgltcaagaacatcgttgaagttccacatgaagctgaagccaa 536
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 250 CTGCCACACCTGCTGAGACGACGACATGAGGAGGTGAAGCAGCAGGACACACACTGG 309

QY 537 gccaccatcgaatgtaacacagtgtaagtgagccaccgccttgcttcagtactgt 556
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 310 GGGCCCCGTCAACAAGACTGTCACAGCGGACCCAGGCTTCTCTGCTGCTGCTTC 369

QY 597 gccaccagacatggaagctgcagctcgaactgctgataaagctccttcctggtgaac 656
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 370 GCGCCCGTCTGCGTGCACCGGCCCATCTACCCGTGCTGCTGCTGCTGCGAGCCGTCGC 429

QY 657 gcccttagtaagcaggtcatgaactcctcctagtgcatccctgcccaatcagtgaaaac 716
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 430 GACTCGTGGAGCCGCGTCAATGACAGTTCTTGGCTTACTGCGCCGAGATGTTAAGTGT 489

QY 717 cagctgtgtccggtgatgaggtcttc 746
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 490 GACAAAGTCCCGAGGGGAGACGTCTGCATC 519

RESULT 4
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;   APPLICANT: FLEISCHMAN, Robert D.
;   APPLICANT: WHITE, Owen R.
;   APPLICANT: FRASER, Claire M.
;   APPLICANT: VENNEN, John C.
;   TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;   TITLE OF INVENTION: TUBERCULOSIS
;   FILE REFERENCE: 24366-20007.00
;   CURRENT APPLICATION NUMBER: US/09/103,840A
;   CURRENT FILING DATE: 1998-06-24
;   NUMBER OF SEQ ID NOS: 2
;   SOFTWARE: PatentIn Ver. 2.1
;   SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
```

```
;
;   OTHER INFORMATION: represent a, t, c or g
;   US-09-103-840A-2

Query Match
Best Local Similarity 50.0%; Score 38.4; DB 4; Length 4403765;
Matches 96; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 465 gacatggtgctgattcaaacagccctgtgtgaagacatctgtgtgaagttccacatgaag 524
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 291448 GAGGTGTTCGTCGACAGGCGCCACAGTGTGCGCATGATGTGATGCGCGCGAG 291389

QY 525 actgaagcccaagccacatcccgatggaacacaggtgcaagtggtcccaaccgcctgtc 584
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 291388 AACCTGGCCGAAMCCCGCAGCAAGTGGGGGACCGCGCTGTGCTGACGCTCACCGCC 291329

QY 585 ctcaagtactgtgcaacagccatggaagctgagcatccaactgctgataagctcttc 644
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 291328 GACGATGCGCGTCGACAAATCAGCAGACACTCGCGACCAACGATGCGCAAGCCGAC 291269

QY 645 ttctgtgtgaac 656
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 291268 ATCTGTGTCAC 291257

RESULT 5
US-09-130-114-2
; Sequence 2, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
;   APPLICANT: HORLICK, Robert A.
;   APPLICANT: DAMEJ, Bassam B.
;   APPLICANT: ROBINS, Alan K.
;   TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
;   TITLE OF INVENTION: From Multiple Transfected Epithomes
;   FILE REFERENCE: 0867/ID903051
;   CURRENT APPLICATION NUMBER: US/09/130,114
;   CURRENT FILING DATE: 1998-08-06
;   NUMBER OF SEQ ID NOS: 36
;   SOFTWARE: FastSeq for Windows Version 3.0
;   SEQ ID NO 2
;   LENGTH: 1931
;   TYPE: DNA
;   ORGANISM: EBNA
;   US-09-130-114-2

Query Match
Best Local Similarity 43.9%; Score 38.2; DB 2; Length 1931;
Matches 163; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

QY 331 tccctgtgtgagcagctgtgtaacacgctcgtgaagacacatcatctgtgtaaggtca 390
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 624 tcccgctcccccgtctcgtctcccgctcgtctcgtctcgtctcgtctcgtctcgtctc 683

QY 391 tcacagctaacatccctcgaagctggaagcctcccggaatgacacaggaagctgtag 450
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 684 tcccgctcctcgtctcgtctcgtctcgtctcgtctcgtctcgtctcgtctcgtctcgt 743

QY 451 tcaagatcccccgtgaagctgtgctgattcaacacgccccctgtgtaagaagcatcgttg 510
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 744 tcccgctcctcgtctcgtctcgtctcgtctcgtctcgtctcgtctcgtctcgtctcgt 803

QY 511 agttccacatgaagctgaaggtcccaagccacatccgcatgtagacacacagtgaaagtgcc 570
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 804 ccgtctcgtctcgtctcgtctcgtctcgtctcgtctcgtctcgtctcgtctcgtctcgt 863

QY 571 ccaccgcctgtgctcctcagtgactgtgcaacacagcatggaagctgtgcgcatccaatgctc 630
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 864 tgtctcctccgctcctcccgctcgtctcgtctcgtctcgtctcgtctcgtctcgtctcgtc 923

QY 631 tgcataagctctctcctcgtgtggaagcgtttagtaagcaggtgcatgaacctccttagtgc 690
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

QY	Db
691 catccctggccc 701	924 tctcccccgtctactctctctcccggtctcgtccctccctccctcc 983
984 ggcctccagctc 994	

RESULT 6

```

US-08-804-166-7
; Sequence 7, Application US/08804166
; Patent No. 6193972
; GENERAL INFORMATION:
; APPLICANT: Campbell, Robert K.
; APPLICANT: Jameson, Bradford A.
; APPLICANT: Chappel, Scott C.
; TITLE OF INVENTION: HYBRID PROTEINS
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESS: BROWDY AND NEWMARK
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 22207
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,166
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/011,936
; FILING DATE: 20 February 1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: CAMPBELL-2A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SRD ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1301 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 279..1287
; US-08-804-166-7

```

Query Match	2.28;	Score 37.6;	DB 4;	Length 1301;
Best Local Similarity	51.89;	Prod 1.1		

Matches	85	Conservative	0	Mismatches	79	Indels	0	Gaps	0
---------	----	--------------	---	------------	----	--------	---	------	---

OY 283 cagcagcctgcgcctcgtctcagtcgacatgcggaaagccagcgcggagatccctcgtctg 34  
 Db 856 CAGCTGGTGTGGTTCACAGGTCGCCGCCCATCATGTGCACCCCTGGCTGGAGAGAGG 915  
 OY 343 gcagcctcgttaaacacgtcctctgaagcacatcctcgtgaagctatcaagctaca 402  
 Db 916 GCTGCCCCGTGTGCATCACCGTCAACACACCATCTGTGCCGCTACTGCCCCACCATGA 975  
 OY 403 tcttcagcctgcagctgaagccctccggccaatgaacccagagctc 446  
 Db 976 CCCGCGTCTCAGAGGGGCTCTCCGGGCCCTGCGCTCAGAGGTG 1019

RESULT 7  
US-08-910-991-7  
; Sequence 7, Application US/08910991  
; Patent No. 6194177

## RESULT 7

```

US-08-910-991-7
Sequence 7, Application US/08910991
Patent No. 6194177
GENERAL INFORMATION:
APPLICANT: Campbell, Robert K.
APPLICANT: Jameson, Bradford A.
APPLICANT: Chappel, Scott C.
TITLE OF INVENTION: HYBRID PROTEINS
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 22207
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,991
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/804,166
FILING DATE: 20 February 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/011,996
FILING DATE: 20 February 1996
ATTORNEY/AGENT INFORMATION:
NAME: YUN, Allen C.
REGISTRATION NUMBER: 37,971
REFERENCE/DOCKET NUMBER: CAMPBELL=2B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (402) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ. ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1301 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 279..1287
US-08-910-991-7

```

Query Match	2.28;	Score 37.6;	DB 4;	Length 1301;
-------------	-------	-------------	-------	--------------

Best local similarity	51.8%;	Pred. No. 1.1;							
Matches	85;	Conservative	0;	Mismatches	79;	Indels	0;	Gaps	0;

Qy	283	cagcagctgycgcgtgctcattgtccatgcyggnaaaagccagccggagagatccctgtgctg	342
Db	856	CACTCTGTTGTTGTTCCAGGCTGCGGCCCATCATGCAACCTGGGTGTGGAGAAAGGAGG	915
Qy	343	gagagctgtgtgaaccgctcctgtgaagaacatcatctgtgctgtgaagttcatcagaattaca	402
Db	916	GGGGCCCCCTGGGATCACCTTCACACCCACCACTGTCTCCGGTAACTGCCCCACCATGA	975
Qy	403	tctctcagctcaggttgaaagccctcgggccaatgtacaggaagtgt	446
Db	976	CCGCGTCTGTGAGGGGGTCTCTGCCGGCCCTCGTCAAGTGTGTG	1019

RESULT 8  
ITS-09-434-288-11

; Sequence 11, Application US/09434288



```
; Patent No. 6303767
; GENERAL INFORMATION:
; APPLICANT: Belach C., Melanie
; APPLICANT: McDaniel, Robert
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
; FILE REFERENCE: 30062-20030.00
; CURRENT APPLICATION NUMBER: US/09/434,288
; CURRENT FILING DATE: 1999-11-05
; PRIOR APPLICATION NUMBER: 60/107,093
; PRIOR FILING DATE: 1998-11-05
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 11
; LENGTH: 3241
; TYPE: DNA
; ORGANISM: Streptomyces narbonensis
; US-09-434-288-11
```

```
Query Match          2.2%; Score 37.4; DB 4; Length 3241;
Best Local Similarity 47.3%; Pred.No.1.9; Indels 0; Gaps 0;
Matches 113; Conservative 0; Mismatches 126;
```

```
QY 319 ccagccgagagatccctgtgctggtgagcagctgtgacacccgtcctgaagcacatcatct 378
   || || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1544 cctgcgcgcaagcagcagcgtgtcaacaacgctgtgcacagctgtgggtcttccaggct 1603
   || || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 379 gactgaagtcatacagcagcattccatccatcagctgcagctgaagccctggccaatgacc 438
   || || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1604 gggtagtgcctgactgtgtccaccgccggcagcagcagccatcaccaagggtcctcgacc 1663
   || || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 439 aagagctgtgactcaagatcccccctgtgacatggtgtgctgattcaacacgcccctgtga 498
   || || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1664 aggaagctggcgtcgagctcccccgcgacatcccgcgagcgagccctgcgcgcgcca 1723
   || || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 499 agaccatcgttgagttccacatgacgactgagggcccaagccacatccgcatgagacac 557
   || || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1724 agttctcggtgacgcgtcgaagcagcgcgtctctgaacgcgacaggtcccgagcgagcc 1782
   || || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

## RESULT 9

```
US-08-471-033-39
; Sequence 39, Application US/08471033
; Patent No. 5770696
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory M
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalin M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,033
; FILING DATE:
; CLASSIFICATION: 530
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: P-40,403
; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SOLV3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1241 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Synthetic DNA"
; HYPOTHEICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 9..1238
; OTHER INFORMATION: /note="Maize optimized DNA
; OTHER INFORMATION: sequence encoding VIP2a(a) with the Bacillus secretion sig
; OTHER INFORMATION: removed as contained in pcIB5527"
; US-08-471-033-39
```

```
Query Match          2.2%; Score 37.2; DB 1; Length 1241;
Best Local Similarity 48.6%; Pred.No.1.3;
Matches 102; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
```

```
QY 353 gaaccagctcctgaagcaatcatctgctgtagatcatcagctcaaatcctcagct 412
   || || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 155 GAACGACATCAAGACCACTACAGAGAGATCACCCTTCAGATGACCGCAGCTTCAGGA 214
   || || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 413 gcaggtgaagccctcgccaatgacagagctgtctagtcgaatccctggacatggt 472
   || || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 215 CGAGATCAAGAGACCTGAGAGATGCACAAGATGTTGACACACCACTGAGCAACG 274
   || || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 473 gctgattcaacagccctgtgtaagaccatcgtgagttccacatgagcactgagc 532
   || || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 275 CATCATCACCCTACACAGAGGTGAGGCCACACCATCGCTTCACACAGAGCCTGACC 334
   || || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 533 ccaagccaccatccgcatgagaccagctgc 562
   || || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 335 GGGCAACACCATCAACAGCGACGCGCATGTC 364
   || || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

## RESULT 10

```
US-08-471-033-42
; Sequence 42, Application US/08471033
; Patent No. 5770696
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory M
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalin M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
```

Query Match	2.2%	Score 37.2;	DB 1;	Length 1241;
Best Local Similarity	48.6%;	Pred. No. 1.3;	Mismatches 108;	Indels 0;
Matches	102;	Conservative	0;	Gaps 0;
QY	353	gaacacgctccatgaagacacatcatctggtcgtgaaggtcaltcacagctaaatcctccagct	412	
Db	155	GAACGACATCTCAAGACCACTACCAAGAGATCACTTCACCAATAGCCGGAGCTTCAGGA	214	
QY	413	gcaggtgaagcccttgcgccaatgaccagagctgtctagtcaaagatcccccgtggaatggt	472	
Db	215	CGAGTCTCAAGGACCTGGAAGGAAATGACAAAGATGTTGCAAGAGCAACCTGAGCAACAG	274	
QY	473	gactgattacacaagcccccttggtcgaagaccactgtgagttgcacatgagacatgagc	532	
Db	275	CATCATCACTACCAAGAAAGAGCTGGAGCCCAACCAACCTCGGCTTCAACAAAGCCTGACGA	334	
QY	533	ccaagccacccatccgcattggaacaccagtgc	562	
Db	335	GGGCAACACCATCAACAGGACGAGCGCATGCG	364	

RESULT 11  
US-08-471-044-39  
; Sequence 39, Application US/08471044

```

1 Patent No. 5840868
2 GENERAL INFORMATION:
3 APPLICANT: Warren, Gregory W
4 APPLICANT: Koziel, Michael G
5 APPLICANT: Mullins, Martha A
6 APPLICANT: Nye, Gordon J
7 APPLICANT: Carr, Brian
8 APPLICANT: Desai, Nalini M
9 APPLICANT: Kostelchka, N. Kristy
10 APPLICANT: Duck, Nicholas B
11 APPLICANT: Estruch, Juan J
12 TITLE OF INVENTION: No. 5840868e1 Pesticidal Proteins and Strains
13 NUMBER OF SEQUENCES: 50
14 CORRESPONDENCE ADDRESS:
15 ADDRESSEE: CIBA-GEIGY Corporation
16 STREET: 7 Skyline Drive
17 CITY: Hawthorne
18 STATE: NY
19 COUNTRY: USA
20 ZIP: 10532
21 COMPUTER READABLE FORM:
22 MEDIUM TYPE: Floppy disk
23 COMPUTER: IBM PC compatible
24 OPERATING SYSTEM: PC-DOS/MS-DOS
25 SOFTWARE: Patentin Release #1.0, Version #1.30B
26 CURRENT APPLICATION DATA:
27 APPLICATION NUMBER: US/08/471,044
28 FILING DATE: 06-JUN-1995
29 CLASSIFICATION: 800
30 PRIOR APPLICATION DATA:
31 APPLICATION NUMBER: US 08/463,483
32 FILING DATE: 05-JUN-1995
33 PRIOR APPLICATION DATA:
34 APPLICATION NUMBER: US 08/314,594
35 FILING DATE: 09-SEP-1994
36 PRIOR APPLICATION DATA:
37 APPLICATION NUMBER: US 08/218,018
38 FILING DATE: 23-MAR-1994
39 PRIOR APPLICATION DATA:
40 APPLICATION NUMBER: US 08/037,057
41 FILING DATE: 25-MAR-1993
42 ATTORNEY/AGENT INFORMATION:
43 NAME: Pace, Gary M.
44 REGISTRATION NUMBER: 40,403
45 REFERENCE/DOCKET NUMBER: CGC 1695/CLP3/DIV6 - SOLV3
46 TELECOMMUNICATION INFORMATION:
47 TELEPHONE: 919-541-8582
48 TELEFAX: 919-541-8689
49 INFORMATION FOR SEQ ID NO: 39:
50 SEQUENCE CHARACTERISTICS:
51 LENGTH: 1241 base pairs
52 TYPE: nucleic acid
53 STRANDEDNESS: single
54 TOPOLOGY: linear
55 MOLECULE TYPE: other nucleic acid
56 DESCRIPTION: /desc = "Synthetic DNA"
57 HYPOTHETICAL: NO
58 FEATURE:
59 NAME/KEY: CDS
60 LOCATION: 9..1238
61 OTHER INFORMATION: /note= "maize optimized DNA
62 OTHER INFORMATION: sequence encoding VIP2a(a) with the Bacillus secretion signal
63 OTHER INFORMATION: removed as contained in pcIB5527"
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898

```

Query Match	Similarity	48.6%	Score 37.2	DB 2	Length 1241
Best Local	Similarity	48.6%	Pred. No. 1.3		
Matches 102	Conservative	0	Mismatches 108	Indels 0	Gaps 0
QY	353	gaacaccgctctgaagacacatcatcttgctgctgaagtcataccagctaacatctccagct	412		
Db	155	gaagcgacatcgaagaccacaaatcgaaggaatcacccttgcagctatccggcgcttcgacga	214		



```
ATTORNEY/AGENT INFORMATION:
NAME: Spruill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8615
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 1241 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 9..1238
OTHER INFORMATION: sequence encoding VIP2A(a) with the Bacillus secretion signal
OTHER INFORMATION: removed as contained in pcIB5527"
US-08-463-483A-39
```

```
Query Match
Best Local Similarity 48.6%; Score 37.2; DB 2; Length 1241;
Matches 102; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
```

```
QY 353 gaacacgctcgaagcaccatcatcgtgctgaagtcatacctaaccatccacgt 412
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 155 GAAGCAGATCAAGACCACTACAGAGATCACTTCAGCAGCGGACCTTGAGGA 214
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 413 gcaggtgaagccctcgcccaatgaccagagctgtctgaagatcccccctgacatg 472
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 215 CGAGATCAAGACCTGGAAGAGATCGACAGATGTCGACAAAGACCAACCTGAGCAACAG 274
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 473 ggcgtgattcaacagccctcggtcaagaccatcgtgtgagttccacatgacgtgagc 532
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 275 CATCATCACTTACAAAGAGAGTGGAGCCACCCATCCATCGGCTTCAAGAGCCCTGACCGA 334
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 533 ccaagccaccatccgcatgacacacagtcg 562
    |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 335 GGGCAACACCATCAACAGCGACGCCCATGGC 364
    |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
```

```
RESULT 14
US-08-463-483A-42
Sequence 42, Application US/08463483A
Patent No. 5849870
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Kozziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5849870e1 Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESSES:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,483A
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Spruill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8615
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 1241 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 9..1238
OTHER INFORMATION: sequence encoding VIP2A(a) with the Bacillus secretion sig
OTHER INFORMATION: removed and the eukaryotic secretion signal inserted as
OTHER INFORMATION: contained in pcIB5528"
US-08-463-483A-42
```

```
Query Match
Best Local Similarity 48.6%; Score 37.2; DB 2; Length 1241;
Matches 102; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
```

```
QY 353 gaacacgctcgaagcaccatcatcgtgctgaagtcatacctaaccatccacgt 412
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 155 GAAGCAGATCAAGACCACTACAGAGATCACTTCAGCAGCGGACCTTGAGGA 214
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 413 gcaggtgaagccctcgcccaatgaccagagctgtctgaagatcccccctgacatg 472
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 215 CGAGATCAAGACCTGGAAGAGATCGACAGATGTCGACAAAGACCAACCTGAGCAACAG 274
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 473 ggcgtgattcaacagccctcggtcaagaccatcgtgtgagttccacatgacgtgagc 532
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 275 CATCATCACTTACAAAGAGAGTGGAGCCACCCATCCATCGGCTTCAAGAGCCCTGACCGA 334
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 533 ccaagccaccatccgcatgacacacagtcg 562
    |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 335 GGGCAACACCATCAACAGCGACGCCCATGGC 364
    |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
```

```
RESULT 15
US-08-471-046A-39
Sequence 39, Application US/08471046A
Patent No. 5866326
GENERAL INFORMATION:
```

```
APPLICANT: Warren, Gregory W
APPLICANT: Kozziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
```

APPLICANT: Duck, Nicholas B  
APPLICANT: Estruch, Juan J  
TITLE OF INVENTION: Method for Isolating Vegetative Insecticidal  
TITLE OF INVENTION: Protein Genes  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 586632artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: NC  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/471,046A  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/463,483  
FILING DATE: 05-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/314,594  
FILING DATE: 09-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/218,018  
FILING DATE: 23-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/037,057  
FILING DATE: 25-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Weigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV8 - SOLV4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 39:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1241 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "Synthetic DNA"  
HYPOTHEICAL: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 9..1238  
OTHER INFORMATION: /note= "Maize optimized DNA  
OTHER INFORMATION: sequence encoding VIP2(a) with the Bacillus secretion signal  
OTHER INFORMATION: removed as contained in pc1b5527"  
US-08-471-046A-39

Query Match 2.2%; Score 37.2; DB 2; Length 1241;  
Best Local Similarity 48.6%; Pred. No. 1.3;  
Matches 102; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 353 gaacacgctcctgaagacacatcatgtgtgaaggtcattcacagctaacatctccagct 412  
DB 155 GAACGACATCAAGACCACTACAGAGATCACTTCAGCATAGCCGCGACGCTTCGAGGA 214  
QY 413 gaagtgaaagccctggyccaatgacagagagctgctagtcgaagctcccttgagacatg 472  
DB 215 CGAGATCAAGACCTTAAGAGATCGACAAGATGTTGACACAGACCAACTTGAAGCAGAG 274  
QY 473 ggcctgattcaacacgcccctgtgccaagacatcgltgaggtccacatgacgactgagc 532  
DB 275 CATCATCACCTTACAGAGAGCTGAGGCCACACCATCGGCTTCAACAGAGCGCTGACGGA 334

QY 533 ccaagccaccatccgcatgtgacacccagtcg 562  
DB 335 GGGCAACACCATCAACAGCGGACGCCATGCGC 364

Search completed: September 16, 2002, 03:26:41  
Job time: 9324 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 06:31:11 : Search time 390.66 Seconds  
(without alignments)  
1384.396 Million cell updates/sec

Title: US-09-700-770-4

Perfect score: 315  
Sequence: 1 taacactgactcagattt.....tgtgtattactactatgc 315

Scoring table:  
OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1736436 segs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N.Geneseq\_032802:\*

1: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	315	100.0	315	21	AAZ29725
2	21	6.7	133894	17	AAT13635
3	18	5.7	9620	18	AAT45424
4	18	5.7	23934	22	ABA19145
5	18	5.7	23934	22	AAL36171
6	18	5.7	23934	22	AAL36179
7	18	5.7	23934	22	AAL04522
8	18	5.7	23934	22	AAS28343
9	18	5.7	23934	22	AAK71442
					Human Lung specific
					ACNPV genomic DNA
					Murine Ataxia-tela
					Human nervous syst
					Human musculoskele
					Human musculoskele
					Human reproductive
					Genomic sequence #
					Human immune/haema

10	17	5.4	699	22	AAH92576	Human inflammatory
11	17	5.4	726	22	AAH04423	Human cDNA clone (
12	17	5.4	842	21	AAZ36779	Nucleotide sequenc
13	17	5.4	1239	22	AAH94373	Human foetal cDNA,
14	17	5.4	1959	22	ABA45382	Human breast cell
15	17	5.4	1959	22	ABA55872	Human foetal liver
16	17	5.4	1959	22	ABA25546	Probe #4012 for ge
17	17	5.4	1959	22	AAK04090	Human brain expres
18	17	5.4	1959	22	AAK29571	Human bone marrow
19	17	5.4	1959	22	AAI14146	Probe #4079 for ge
20	17	5.4	1959	22	AAI35531	Probe #4217 used t
21	17	5.4	1959	22	AAI03994	Arabidopsis thalia
22	17	5.4	2388	21	AAC36512	Arabidopsis thalia
23	17	5.4	2879	23	ABL27076	Drosophila melanog
24	17	5.4	3008	23	ABL21896	Drosophila melanog
25	17	5.4	3201	23	ABL02390	Drosophila melanog
26	17	5.4	3241	23	ABL20736	Drosophila melanog
27	17	5.4	3424	23	ABL21898	Drosophila melanog
28	17	5.4	4060	23	ABL07105	Arabidopsis thalia
29	17	5.4	4146	17	AAT09026	Human cardiovascular
30	17	5.4	4645	22	AAS35982	Human cardiovascular
31	17	5.4	4646	22	AAS35981	Human cardiovascular
32	17	5.4	5406	22	AAF83912	cDNA sequence of p
33	17	5.4	5407	23	ABL18530	Drosophila melanog
34	17	5.4	5408	21	AAA10226	Human PCNA-1 splic
35	17	5.4	5532	22	AAF83913	cDNA sequence of p
36	17	5.4	5534	21	AAAI0227	Human PCNA-1 splic
37	17	5.4	6236	23	ABL23206	Drosophila melanog
38	17	5.4	6236	23	ABL20966	Drosophila melanog
39	17	5.4	9931	23	ABL07088	Drosophila melanog
40	17	5.4	12151	21	AAZ36757	DNA sequence encod
41	17	5.4	15866	22	AAK90892	Human digestive sy
42	17	5.4	15866	22	AAK90892	Human digestive sy
43	17	5.4	15866	22	AAK90892	Human digestive sy
44	17	5.4	15866	22	AAK90892	Human digestive sy
45	17	5.4	21403	22	AAI04555	Human reproductive

ALIGNMENTS

RESULT 1	
AAZ29725	standard; DNA; 315 BP.
AC	AAZ29725;
AC	AAZ29725;
DT	27-MAR-2000 (first entry)
XX	Human Lung specific gene-2.
DE	Human Lung specific gene-2.
XX	Lung Specific Gene; LSG; human; diagnostic marker;
KW	prognosticate; lung cancer; diagnosis; ds.
XX	
OS	Homo sapiens.
XX	
PN	W09960160-A1.
XX	
PD	25-NOV-1999.
XX	
PF	12-MAY-1999; 99WO-US10344.
XX	
PR	21-MAY-1998; 98US-0086212.
XX	
PA	(DIAD-) DIADEXUS LLC.
XX	
PI	Yang F, Macina RA, Sun Y;
XX	
DR	WPI; 2000-116320/10.
XX	
PT	A new method for diagnosing, monitoring and staging lung cancer
XX	
PS	Example 1; Page 35; 40pp; English.

```

XX CC The present sequence is a lung specific gene (LSG) from human
CC CC clone ID 1472038. The LSG has high level of tissue specificity for lungs
CC CC and is overexpressed in cancerous tissues. The sequence serves as a
CC CC diagnostic marker for detecting, monitoring, staging and prognosticating
CC CC lung cancer. The diagnosis involves comparing levels of LSG in samples
CC CC obtained from patient and normal control.
XX SQ Sequence 315 BP; 108 A; 58 C; 49 G; 100 T; 0 other;

Query Match
Best Local Similarity 100.0%; Score 315; DB 21; Length 315;
Matches 315; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taaacactgactcagatttaagaataactttgagaataagaacaatgaatcagtt 60
Db 1 taaacactgactcagatttaagaataactttgagaataagaacaatgaatcagtt 60
QY 61 tctccacacttaagtatactcttaagatctacagctccctccttaaggagacatacaa 120
Db 61 tctccacacttaagtatactcttaagatctacagctccctccttaaggagacatacaa 120
QY 121 gtccagttgtgtccttctgttgatccacccttatatcaagtagtgatgactacaatt 180
Db 121 gtccagttgtgtccttctgttgatccacccttatatcaagtagtgatgactacaatt 180
QY 181 ttgaataatagttgcacacataaactgtgagttatgagaacacatgtaagaagaata 240
Db 181 ttgaataatagttgcacacataaactgtgagttatgagaacacatgtaagaagaata 240
QY 241 caacattccaccccttaacagagatcattacttgcacacacgaataattgtcatgtt 300
Db 241 caacattccaccccttaacagagatcattacttgcacacacgaataattgtcatgtt 300
QY 301 attactactatgc 315
Db 301 attactactatgc 315

RESULT 2
AAT13635/c
ID AAT13635 standard; DNA: 133894 BP.
XX AC AAT13635;
XX DT 03-SEP-1996 (first entry)
XX DE ACNPV genomic DNA clone 6.
XX KM Autographa californica nuclear polyhedrosis virus clone 6;
XX KM disruption; non-essential gene; heterologous protein production;
XX KM expression vector; baculovirus; ss.
XX OS Autographa californica nuclear polyhedrosis virus clone 6.
XX PN WO9601320-A2.
XX PD 18-JAN-1996.
XX PF 30-JUN-1995; 95WO-IB00578.
XX PR 04-JUL-1994; 94GB-0013420.
XX RA (NATU-) NATURAL ENVIRONMENT RES COUNCIL.
XX PI Ayres M, Bishop D, Possee R;
XX DR WPI: 1996-087670/09.
XX DR GENBANK: L22858.
XX PT Autographa californica nuclear polyhedrosis virus complete genome
PT sequence - useful in the prodn. of vectors for enhanced

```

```

PT heterologous protein expression, such as interleukin(s),
PT interferon(s) and neurotoxin(s)
XX PS Disclosure; Page 90-186; 122pp; English.
XX CC The complete nucleotide sequence of the genome of clone 6 of the
CC CC baculovirus Autographa californica nuclear polyhedrosis virus (AcNPV)
CC CC has been determined. The sequence is taken from the Genbank record
CC CC L22858. The patent specification claims a polynucleotide selected from
CC CC open reading frames (ORFs 13, 20, 22-26, 28-30, 32, 38, 41-46, 50-60,
CC CC 62-63, 66, 68-79, 81-87, 91-92, 96-98, 101-103, 106-126, 129-130,
CC CC 140-146, 148-150, 152 and 154 from a total of 154 ORFs identified by
CC CC the patentees. See T13636-731. Expression vectors contg. the complete
CC CC genomic sequence of AcNPV, with the exception that at least one non-
CC CC essential ORF is disrupted or replaced are useful for the synthesis of
CC CC heterologous proteins.
XX SQ Sequence 133894 BP; 39195 A; 27151 C; 27347 G; 40201 T; 0 other;

Query Match
Best Local Similarity 6.7%; Score 21; DB 17; Length 133894;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 aagaataactttgagaat 41
Db 114255 AACAAATTAACCTTTGAGAAAT 114235

RESULT 3
AAT45424
ID AAT45424 standard; cDNA: 9620 BP.
XX AC AAT45424;
XX DT 30-MAR-1998 (first entry)
XX DE Murine Ataxia-telangiectasia ATM gene.
XX KM Ataxia-telangiectasia; A-T; mutated; ATM; 11q22-23; signal transduction;
XX KM DNA damage; cell cycle control; screening; gene therapy; catalytic;
XX KM Phosphatidylinositol-3 kinase; PI-3; cancer predisposition; carrier;
XX KM mouse; ss.
XX OS Mus musculus.
XX FH Key Location/Qualifiers
XX FT CDS 42..9242
XX FT /*tag= a
XX PN WO9636695-A1.
XX PD 21-NOV-1996.
XX PF 16-MAY-1996; 96WO-US07040.
XX PR 28-JUL-1995; 95US-0508836.
XX PR 16-MAY-1995; 95US-0441822.
XX PR 21-JUN-1995; 95US-0493092.
XX RA (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
XX RA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX PI Collins FS, Shiloh Y, Tagle DA;
XX DR P-PSDB; AAM36178.
XX PT New gene ATM implicated in ataxia-telangiectasia and related protein
XX PT carriers
XX PT - useful in screening methods, partic. for identifying disease
XX PS Example 4; Page 90-95; 127pp; English.

```



XX This sequence represents the mouse ATM gene, mutations in which cause  
CC ataxia-telangiectasia (A-T), a progressive genetic disorder affecting  
CC the central nervous and immune systems. In the human ATM gene, there is  
CC extensive alternate splicing at the 5' UTR of the ATM transcript giving  
CC rise to 12 different 5'UTRs (the longest is given in AAT45421). The  
CC first exon in the human sequence is designated 1b. An alternative leader  
CC exon, 1a is given in AAT45422 and the 3' UTR is in AAT45420. The ATM  
CC gene, located in humans at chromosome 11q22-23, is probably involved in a  
CC novel signal transduction system that links DNA damage surveillance to  
CC cell cycle control. This murine sequence is located in chromosome 9, band  
CC 9C. The human ATM gene product (AAM07655) has a highly conserved  
CC C-terminal region showing high sequence homology to the catalytic domain  
CC of phosphatidylinositol-3 kinases. A-T mutations affect a variety of  
CC tissues and lead to cancer predisposition. Identification of A-T  
CC carriers, by analysis at nucleic acid or protein levels, allows better  
CC supervision and treatment of such subjects who are at increased risk of  
CC developing cancer and are particularly sensitive to radiation. The  
CC transgenic animals and transformed cells are useful as models of the  
CC human disease. Also viral vectors expressing the ATM protein can be used  
CC in gene therapy of A-T.

XX  
SQ Sequence 9620 BP; 2898 A; 1931 C; 2174 G; 2617 T; 0 other;

Query Match 5.7%; Score 18; DB 18; Length 9620;

Best Local Similarity 100.0%; Pred. No. 22;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 80 tctcttagagatctacag 97  
|||||

Db 6071 tctcttagagatctacag 6088

#### RESULT 4

ABAI9145

ID ABA19145 standard; DNA; 23934 BP.

XX ABA19145;

DT 23-JAN-2002 (first entry)

DE Human nervous system related polynucleotide SEQ ID NO 11476.

XX Human; nocitropic; neuroprotective; cytostatic; dermatological; virocidic;

KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;

KW antiparkinsonian; antistickling; antianaemic; antiarthritic; cancer;

KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;

KW antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;

KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;

KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.

XX Homo sapiens.

PN WO200159063-A2.

PD 16-AUG-2001.

PF 17-JAN-2001; 2001WO-US01334.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0188874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 23-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232339.  
PR 14-SEP-2000; 2000US-0232396.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239933.  
PR 13-OCT-2000; 2000US-0239935.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.

PR 20-OCT-2000; 2000US-0242221.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 17-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249284.  
PR 17-NOV-2000; 2000US-0249285.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 01-DEC-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250391.  
PR 01-DEC-2000; 2000US-0251160.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251388.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251858.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0251990.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM.  
PI  
DR WPI; 2001-541565/60.  
XX  
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating nervous system  
PT cancers and metastases -  
XX  
PS Disclosure; SEQ ID NO 11476; 1701bp + Sequence Listing; English.  
XX  
XX The invention relates to novel genes (ABR11004-ABR21534) and proteins  
CC (ABR14678-ABR18001) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's

CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischemias;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pt\_sequences.  
XX  
SQ Sequence 23934 BP; 7785 A; 4761 C; 4365 G; 7023 T; 0 other;  
  
Query Match 5.7%; Score 18; DB 22; Length 23934;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 172 ctacaatttgaata 189  
DB 10877 ctacaatttgaata 10894  
|||||  
  
RESULT 5  
AAL36171/C  
ID AAL36171 standard; DNA; 23934 BP.  
XX  
AC AAL36171;  
XX  
DT 08-JAN-2002 (first entry)  
XX  
DE Human musculoskeletal system related polynucleotide SEQ ID NO 2536.  
XX  
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; anticancer;  
KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW caritant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein;  
KW musculoskeletal system; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO20015367-A1.  
XX  
PD 02-AUG-2001.  
XX  
XX  
PF 17-JAN-2001; 2001WO-US01338.  
XX  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.



Query Match 5.7%; Score 18; DB 22; Length 23934;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 172 ctacaatttggaaaata 189  
|||||  
Db 13058 CTACAAATTTGAAAATA 13041

RESULT 6  
AAL36179  
ID AAL36179 standard; DNA: 23934 BP.  
XX  
AC AAL36179;  
XX  
DT 08-JAN-2002 (first entry)  
XX

Human musculoskeletal system related polynucleotide SEQ ID NO 2544.

KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antilicer;  
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiact; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein;  
KW musculoskeletal system; ds.

OS Homo sapiens.

PN WO200155367-A1.

PD 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US0138.

XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 23-AUG-2000; 2000US-0227182.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0233297.  
PR 14-SEP-2000; 2000US-0233298.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0241826.  
PR 08-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.

PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249219.  
PR 17-NOV-2000; 2000US-0249220.  
PR 17-NOV-2000; 2000US-0249221.  
PR 17-NOV-2000; 2000US-0249222.  
PR 17-NOV-2000; 2000US-0249223.  
PR 17-NOV-2000; 2000US-0249224.  
PR 17-NOV-2000; 2000US-0249225.  
PR 17-NOV-2000; 2000US-0249226.  
PR 17-NOV-2000; 2000US-0249227.  
PR 17-NOV-2000; 2000US-0249228.  
PR 17-NOV-2000; 2000US-0249229.  
PR 17-NOV-2000; 2000US-0249230.  
PR 17-NOV-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251088.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0251990.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-451937/48.  
XX  
XX Isolated polypeptide for treating, preventing and/or prognosing  
XX disorders related to the musculoskeletal system including  
XX musculoskeletal cancers and also for testing and detection e.g.  
XX diagnosis -  
XX  
XX Example 2; SEQ ID NO 254; 781bp + Sequence Listing; English.  
XX  
XX The invention relates to novel genes (AAL04522; AAL04523) and proteins  
XX (AAB03087; AAB04109) associated with the musculoskeletal system useful  
XX for preventing, treating or ameliorating medical conditions e.g. by  
XX protein or gene therapy. The genes are isolated from a range of human  
XX tissues disclosed in the specification. The nucleic acids, proteins,  
XX antibodies and (ant)agonists are useful in the diagnosis, treatment  
XX and prevention of: (a) cancer, e.g. breast and ovarian cancer and  
XX other cancers of the adrenal gland, bone, bone marrow, breast,  
XX gastrointestinal tract, liver, lung, or urogenital; (b) immune  
XX disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
XX anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
XX multiple sclerosis, rheumatoid arthritis and ulcerative colitis;  
XX (c) cardiovascular disorders such as myocardial ischaemias; (d) wound  
XX healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;  
XX and (f) infectious diseases such as viral, bacterial, fungal and  
XX parasitic infections.  
XX Note: The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 23934 BP; 7785 A; 4761 C; 4365 G; 7023 T; 0 other;  
XX  
XX Query Match 5.7%; Score 18; DB 22; Length 23934;  
XX Best Local Similarity 100.0%; Pred. No. 22;  
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX  
XX AAL04522;  
XX  
XX 21-NOV-2001 (first entry)  
XX  
XX Human reproductive system related antigen DNA SEQ ID NO: 7210.  
XX  
XX Human: reproductive system related antigen; reproductive system disorder;  
XX cancer; gene therapy; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO20015320-A2.  
XX  
XX 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US01339.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
XX 04-FEB-2000; 2000US-0180628.  
XX 24-FEB-2000; 2000US-0184664.  
XX 02-MAR-2000; 2000US-0186350.  
XX 16-MAR-2000; 2000US-0189874.  
XX 17-MAR-2000; 2000US-0190076.  
XX 18-APR-2000; 2000US-0198123.  
XX 19-MAY-2000; 2000US-0205515.  
XX 07-JUN-2000; 2000US-0209467.  
XX 28-JUN-2000; 2000US-0214886.  
XX 30-JUN-2000; 2000US-0215135.  
XX 07-JUL-2000; 2000US-0216647.  
XX 11-JUL-2000; 2000US-0216880.  
XX 11-JUL-2000; 2000US-0217487.  
XX 14-JUL-2000; 2000US-0217496.  
XX 26-JUL-2000; 2000US-0218290.  
XX 26-JUL-2000; 2000US-0220963.  
XX 14-AUG-2000; 2000US-0224518.  
XX 14-AUG-2000; 2000US-0224519.  
XX 14-AUG-2000; 2000US-0225213.  
XX 14-AUG-2000; 2000US-0225214.  
XX 14-AUG-2000; 2000US-0225266.  
XX 14-AUG-2000; 2000US-0225267.  
XX 14-AUG-2000; 2000US-0225268.  
XX 14-AUG-2000; 2000US-0225270.  
XX 14-AUG-2000; 2000US-0225447.  
XX 14-AUG-2000; 2000US-0225757.  
XX 14-AUG-2000; 2000US-0225758.  
XX 14-AUG-2000; 2000US-0225759.  
XX 18-AUG-2000; 2000US-0226279.  
XX 22-AUG-2000; 2000US-0226681.  
XX 22-AUG-2000; 2000US-0226686.  
XX 22-AUG-2000; 2000US-0227182.  
XX 23-AUG-2000; 2000US-0227009.  
XX 30-AUG-2000; 2000US-0228924.  
XX 01-SEP-2000; 2000US-0229287.  
XX 01-SEP-2000; 2000US-0229343.  
XX 01-SEP-2000; 2000US-0229344.  
XX 01-SEP-2000; 2000US-0229345.  
XX 05-SEP-2000; 2000US-0229509.  
XX 05-SEP-2000; 2000US-0229513.  
XX 06-SEP-2000; 2000US-0230437.  
XX 06-SEP-2000; 2000US-0230438.  
XX 08-SEP-2000; 2000US-0231242.  
XX 08-SEP-2000; 2000US-0231243.  
XX 08-SEP-2000; 2000US-0231244.  
XX 08-SEP-2000; 2000US-0231245.  
XX 08-SEP-2000; 2000US-0231413.  
XX 08-SEP-2000; 2000US-0231414.  
XX 08-SEP-2000; 2000US-0232080.  
XX 08-SEP-2000; 2000US-0232081.  
XX 12-SEP-2000; 2000US-0231966.  
XX 14-SEP-2000; 2000US-0232397.  
XX 14-SEP-2000; 2000US-0232398.  
XX 14-SEP-2000; 2000US-0232399.

```
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239335.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.

PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM.
XX
XX WPI, 2001-465570/50.
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen
XX is used in preventing, treating or ameliorating a medical condition -
XX
XX Disclosure; SEQ ID NO 7210; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
SO Sequence 23934 BP; 7023 A; 4365 C; 4761 G; 7785 T; 0 other;

Query Match 5.7%; Score 18; DB 22; Length 23934;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 ctcaaatgtgaataaata 189
DB 13058 CTACCAATTTGAAATA 13041

RESULT 8
AAS28343
ID AAS28343 standard; DNA; 23934 BP.
XX
XX AAS28343;
AC
XX
XX 07-NOV-2001 (first entry)
DT
XX
XX Genomic sequence #183 encoding for novel human respiratory antigen.
DE
XX
XX Human; respiratory antigen; respiratory disorder; throat disorder;
KW lung disorder; nose disorder; lung cancer; gene therapy; cytostatic;
KW anti allergic; anti asthmatic; anti inflammatory; olfactory;
KW respiratory active; ds.
XX
XX Homo sapiens.
XX
XX WO20015448-A1.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01333.
XX
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
```

PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 14-AUG-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227182.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 08-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0233397.  
PR 14-SEP-2000; 2000US-0233398.  
PR 14-SEP-2000; 2000US-0233399.  
PR 14-SEP-2000; 2000US-0233400.  
PR 14-SEP-2000; 2000US-0233401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.

PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
  
(HUMA - ) HUMAN GENOME SCI INC.  
Rosen CA, Barash SC, Ruben SM;  
WPI; 2001-476224/51.  
  
Isolated polypeptide for treating, preventing and/ or prognosing  
disorders related to the respiratory system including respiratory  
cancers and also for testing and detection e.g. diagnosis -  
XX  
PS Disclosure; SED ID No 777; 546pp; English.  
  
The present invention relates to the isolation of novel human  
respiratory antigens (AAU17685-AAU17975), and cDNA and genomic  
sequences encoding for these polypeptides. The sequences of the  
invention are useful for preventing, treating and/or prognosing  
disorders related to the respiratory system including throat  
disorders (e.g. vocal cord paralysis, tonsillitis, and laryngitis),  
lung disorders e.g. pneumonia, allergic disorders e.g. asthma,  
CC pleurisy, cystic fibrosis, emphysema, nose disorders and cancers of  
the respiratory tissues e.g. lung cancer. The polynucleotide sequences

CC of the invention are useful in gene therapy and antisense therapy.  
CC AAS28161-AAS28764 represent genomic sequences encoding for novel  
CC human respiratory antigens.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 23934 BP; 7785 A; 4761 C; 4365 G; 7023 T; 0 other;

Query Match 5.7%; Score 18; DB 22; Length 23934;  
Best Local Similarity 100.0%; Pred.No. 22;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 172 ctacaatttgaata 189  
Db 10877 ctacaatttgaata 10894  
|||||

RESULT 9  
AAK71442/c  
ID AAK71442 standard; DNA; 23934 BP.  
XX  
AC AAK71442;  
XX

DT 06-NOV-2001 (first entry)  
XX

DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:26254.  
XX

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ds.  
XX

OS Homo sapiens.  
XX

PN WO200157182-A2.  
XX

PD 09-AUG-2001.  
XX

PF 17-JAN-2001; 2001WO-US01354.  
XX

PR 31-JAN-2000; 2000US-0179065.  
PR

PR 04-FEB-2000; 2000US-0180628.  
PR

PR 24-FEB-2000; 2000US-0184664.  
PR

PR 02-MAR-2000; 2000US-0186350.  
PR

PR 16-MAR-2000; 2000US-0189874.  
PR

PR 17-MAR-2000; 2000US-0190076.  
PR

PR 18-APR-2000; 2000US-0198123.  
PR

PR 19-MAY-2000; 2000US-0205515.  
PR

PR 07-JUN-2000; 2000US-0209467.  
PR

PR 28-JUN-2000; 2000US-0214886.  
PR

PR 30-JUN-2000; 2000US-0215135.  
PR

PR 07-JUL-2000; 2000US-0216647.  
PR

PR 11-JUL-2000; 2000US-0216880.  
PR

PR 11-JUL-2000; 2000US-0217487.  
PR

PR 14-JUL-2000; 2000US-0217496.  
PR

PR 26-JUL-2000; 2000US-0220963.  
PR

PR 26-JUL-2000; 2000US-0220964.  
PR

PR 14-AUG-2000; 2000US-0224518.  
PR

PR 14-AUG-2000; 2000US-0224519.  
PR

PR 14-AUG-2000; 2000US-0225213.  
PR

PR 14-AUG-2000; 2000US-0225214.  
PR

PR 14-AUG-2000; 2000US-0225266.  
PR

PR 14-AUG-2000; 2000US-0225267.  
PR

PR 14-AUG-2000; 2000US-0225268.  
PR

PR 14-AUG-2000; 2000US-0225270.  
PR

PR 14-AUG-2000; 2000US-0225447.  
PR

PR 14-AUG-2000; 2000US-0225757.  
PR

PR 14-AUG-2000; 2000US-0225758.  
PR

PR 14-AUG-2000; 2000US-0225759.  
PR

PR 18-AUG-2000; 2000US-0226279.  
PR

PR 22-AUG-2000; 2000US-0226681.  
PR

PR 22-AUG-2000; 2000US-0226868.  
PR

PR 22-AUG-2000; 2000US-0227182.  
PR

PR 23-AUG-2000; 2000US-0227009.  
PR

PR 30-AUG-2000; 2000US-0228924.  
PR

PR 01-SEP-2000; 2000US-0229287.  
PR

PR 01-SEP-2000; 2000US-0229343.  
PR

PR 01-SEP-2000; 2000US-0229344.  
PR

PR 01-SEP-2000; 2000US-0229345.  
PR

PR 05-SEP-2000; 2000US-0229509.  
PR

PR 05-SEP-2000; 2000US-0229513.  
PR

PR 06-SEP-2000; 2000US-0230437.  
PR

PR 06-SEP-2000; 2000US-0230438.  
PR

PR 08-SEP-2000; 2000US-0231242.  
PR

PR 08-SEP-2000; 2000US-0231243.  
PR

PR 08-SEP-2000; 2000US-0231244.  
PR

PR 08-SEP-2000; 2000US-0231413.  
PR

PR 08-SEP-2000; 2000US-0231414.  
PR

PR 08-SEP-2000; 2000US-0232080.  
PR

PR 08-SEP-2000; 2000US-0232081.  
PR

PR 12-SEP-2000; 2000US-0231968.  
PR

PR 14-SEP-2000; 2000US-0232397.  
PR

PR 14-SEP-2000; 2000US-0232398.  
PR

PR 14-SEP-2000; 2000US-0232399.  
PR

PR 14-SEP-2000; 2000US-0232400.  
PR

PR 14-SEP-2000; 2000US-0232401.  
PR

PR 14-SEP-2000; 2000US-0233063.  
PR

PR 14-SEP-2000; 2000US-0233064.  
PR

PR 14-SEP-2000; 2000US-0233065.  
PR

PR 21-SEP-2000; 2000US-0234223.  
PR

PR 25-SEP-2000; 2000US-0234274.  
PR

PR 25-SEP-2000; 2000US-0234997.  
PR

PR 25-SEP-2000; 2000US-0234998.  
PR

PR 26-SEP-2000; 2000US-0235484.  
PR

PR 27-SEP-2000; 2000US-0235634.  
PR

PR 27-SEP-2000; 2000US-0235636.  
PR

PR 29-SEP-2000; 2000US-0236327.  
PR

PR 29-SEP-2000; 2000US-0236367.  
PR

PR 29-SEP-2000; 2000US-0236368.  
PR

PR 29-SEP-2000; 2000US-0236369.  
PR

PR 29-SEP-2000; 2000US-0236370.  
PR

PR 02-OCT-2000; 2000US-0236802.  
PR

PR 02-OCT-2000; 2000US-0237037.  
PR

PR 02-OCT-2000; 2000US-0237038.  
PR

PR 02-OCT-2000; 2000US-0237039.  
PR

PR 02-OCT-2000; 2000US-0237040.  
PR

PR 13-OCT-2000; 2000US-0239935.  
PR

PR 13-OCT-2000; 2000US-0239937.  
PR

PR 20-OCT-2000; 2000US-0240960.  
PR

PR 20-OCT-2000; 2000US-0241221.  
PR

PR 20-OCT-2000; 2000US-0241785.  
PR

PR 20-OCT-2000; 2000US-0241786.  
PR

PR 20-OCT-2000; 2000US-0241787.  
PR

PR 20-OCT-2000; 2000US-0241808.  
PR

PR 20-OCT-2000; 2000US-0241809.  
PR

PR 01-NOV-2000; 2000US-0241826.  
PR

PR 08-NOV-2000; 2000US-0244617.  
PR

PR 08-NOV-2000; 2000US-0246474.  
PR

PR 08-NOV-2000; 2000US-0246475.  
PR

PR 08-NOV-2000; 2000US-0246476.  
PR

PR 08-NOV-2000; 2000US-0246477.  
PR

PR 08-NOV-2000; 2000US-0246478.  
PR

PR 08-NOV-2000; 2000US-0246523.  
PR

PR 08-NOV-2000; 2000US-0246524.  
PR

PR 08-NOV-2000; 2000US-0246525.  
PR

PR 08-NOV-2000; 2000US-0246526.  
PR

PR 08-NOV-2000; 2000US-0246527.  
PR

PR 08-NOV-2000; 2000US-0246528.  
PR

PR 08-NOV-2000; 2000US-0246532.  
PR

PR 08-NOV-2000; 2000US-0246609.  
PR

PR 08-NOV-2000; 2000US-0246610.  
PR

PR 08-NOV-2000; 2000US-0246611.  
PR

PR 08-NOV-2000; 2000US-0246613.  
PR

PR 17-NOV-2000; 2000US-0249207.  
PR

PR 17-NOV-2000; 2000US-0249208.  
PR



PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
DR WPI: 2001-483426/52.  
XX  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and  
PT metastasis -  
XX  
XX  
XX Disclosure: SEQ ID NO 26254; 3071bp + Sequence Listing; English.  
XX  
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patient's own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/hematopoietic-related diseases, especially  
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
CC represent sequences used in the exemplification of the present invention.  
XX  
XX  
SQ Sequence 23934 BP; 7023 A; 4365 C; 4761 G; 7785 T; 0 other;

Query Match 5.7%; Score 18; DB 22; Length 23934;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 172 ctcaaatatttgaataa 189  
DB 13058 CTACAAATTTGAAATA 13041

RESULT 10  
AAH92576  
ID AAH92576 standard; DNA; 699 BP.  
XX  
XX AAH92576;  
AC  
XX  
XX  
DT 09-OCT-2001 (first entry)  
XX  
XX  
DE Human inflammatory bowel disease related gene fragment IGR1276a.  
XX  
XX Human; inflammatory bowel disease; Crohn's disease; ulcerative colitis;  
KW single nucleotide polymorphism; SNP; chromosome 19p13; paternity test;  
KM chromosome 5q31-33; forensic test; gene therapy; ds.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200142511-A2.  
PN  
XX  
XX 14-JUN-2001.  
PD  
XX  
XX 11-DEC-2000; 2000WO-US33632.  
PF  
XX  
XX 10-DEC-1999; 99US-0170257.  
PR  
XX 10-APR-2000; 2000US-0196046.  
XX  
XX  
XX (MHED) WHITEHEAD INST BIOMEDICAL RES.  
PA (ELLI-) ELLIPISIS BIOTHERAPEUTICS CORP.  
XX  
XX Daly M, Hudson TJ, Lander ES, Rioux J, Siminovitch K;  
PI  
XX  
XX WPI: 2001-367874/38.  
DR  
XX  
XX  
PT Testing for the presence of polymorphisms associated with inflammatory  
PT bowel disease, using a hybridization assay -  
XX  
XX  
XX Disclosure: Page 257; 463bp; English.  
XX  
XX  
XX The present invention describes a method for detecting the presence of  
CC polymorphisms associated with inflammatory bowel diseases such as  
CC ulcerative colitis and Crohn's disease. The methods can be used to detect  
CC the presence of genetic polymorphisms associated with inflammatory bowel  
CC disease and correlating their occurrence with disease states. They may be  
CC used in this way for phenotypic correlations, forensics, paternity  
CC testing, medicine and genetic analysis. The present sequence is a gene  
CC containing a polymorphic site described in the exemplification of the  
XX invention.  
XX  
XX  
SQ Sequence 699 BP; 192 A; 142 C; 99 G; 266 T; 0 other;

Query Match 5.4%; Score 17; DB 22; Length 699;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 44 aacaatgaatcagtt 60  
DB 372 aacaatgaatcagtt 388

RESULT 11  
AAH04423  
ID AAH04423 standard; cDNA; 726 BP.  
XX  
XX  
XX AAH04423;  
AC  
XX  
XX 26-JUN-2001 (first entry)  
DT  
XX  
XX Human cDNA clone (5'-primer) SEQ ID NO:1258.  
DE  
XX  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
OS  
XX  
XX Homo sapiens.  
XX

```

PN  EP1074617-A2.
XX
XX  07-FEB-2001.
XX
XX  28-JUL-2000; 2000EP-0116126.
XX
XX  29-JUL-1999; 99JP-0248036.
XX  27-AUG-1999; 99JP-0300253.
XX  11-JAN-2000; 2000JP-0118776.
XX  02-MAY-2000; 2000JP-0183767.
XX  09-JUN-2000; 2000JP-0241899.
XX
XX  (HELI-) HELIX RES INST.
XX
XX  Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI  Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX  WPI: 2001-318749/34.
XX
XX  Primer sets for synthesizing polynucleotides, particularly the 5602
PT  full-length cDNAs defined in the specification, and for the detection
PT  and/or diagnosis of the abnormality of the proteins encoded by the
PT  full-length cDNAs -
XX
XX  Claim 1; SEQ ID 1258; 2537pp + CD ROM; English.
XX
XX  The present invention describes primer sets for synthesizing 5602
XX  full-length cDNAs defined in the specification. Where a primer set
XX  comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX  to the complementary strand of a polynucleotide which comprises one of
XX  the 5602 nucleotide sequences defined in the specification, where the
XX  oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX  of an oligonucleotide comprising a sequence complementary to the
XX  complementary strand of a polynucleotide which comprises a 5'-end
XX  sequence and an oligonucleotide comprising a sequence complementary to a
XX  polynucleotide which comprises a 3'-end sequence, where the
XX  oligonucleotide comprises at least 15 nucleotides, and the combination of
XX  the 5'-end sequence/3'-end sequence is selected from those defined in
XX  the specification. The primer sets can be used in antisense therapy and
XX  in gene therapy. The primers are useful for synthesizing polynucleotides,
XX  particularly full-length cDNAs. The primers are also useful for the
XX  detection and/or diagnosis of the abnormality of the proteins encoded by
XX  the full-length cDNAs. The primers allow obtaining of the full-length
XX  cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX  AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
XX  AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX  represent oligonucleotides, all of which are used in the exemplification
XX  of the present invention.
XX
XX  Sequence 726 BP; 252 A; 109 C; 145 G; 217 T; 3 other.
XX
XX  Query Match          5.4%; Score 17; DB 22; Length 726;
XX  Best Local Similarity 100.0%; Pred. No. 70;
XX  Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX  Oy 175 caaatlttgaataaga 191
XX  Db 173 caaatlttgaataaga 189
XX
XX  RESULT 12
XX  AAZ36779
XX  ID AAZ36779 standard; DNA; 842 BP.
XX
XX  AAZ36779;
XX
XX  13-MAR-2000 (first entry)
XX
XX  Nucleotide sequence of the human AHCP gene clone 3.
XX
XX  Human; AHCP gene; autosomal highly conserved protein; schizophrenia;
XX  neurological disease; genetic predisposition; chromosome 6p23; D6S274;
XX

```

```

KW  D6S285; psychological disease; gene therapy; ss.
XX
XX  Homo sapiens.
XX
XX  WO9957316-A1.
XX
XX  11-NOV-1999.
XX
XX  30-APR-1999; 99WO-IB00846.
XX
XX  30-APR-1998; 98US-0083625.
XX  31-DEC-1998; 98US-0114592.
XX
XX  (INSP ) INST PASTEUR.
XX  (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
XX  Leroy P, Bourgeron T, McElreavey K, Fellous M, Jamain S;
XX
XX  WPI: 2000-086415/07.
XX
XX  New gene encoding autosomal high conserved protein used to diagnose a
PT  genetic predisposition to schizophrenia -
XX
XX  Example 1; Page 66-67; 76pp; English.
XX
XX  AAZ36777-84 represent overlapping clones of the complete human AHCP
XX  (autosomal highly conserved protein) gene. The AHCP gene is linked
XX  to a genetic predisposition to schizophrenia. The gene is located on
XX  chromosome 6p23, between markers D6S274 and D6S285. Several
XX  polymorphisms are found in the AHCP gene. Oligonucleotide probes
XX  derived from the AHCP sequences can be used to screen for patients
XX  having a genetic predisposition for a neurological or psychological
XX  disease, especially schizophrenia. The invention is used to diagnose
XX  a genetic predisposition to schizophrenia, and to treat the disorder
XX  by gene therapy. The invention provides a treatment that is specific to
XX  schizophrenic disorders, without the risk of significant side effects.
XX
XX  Sequence 842 BP; 258 A; 135 C; 158 G; 291 T; 0 other.
XX
XX  Query Match          5.4%; Score 17; DB 21; Length 842;
XX  Best Local Similarity 100.0%; Pred. No. 70;
XX  Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX  Oy 15 gattttaagaataact 31
XX  Db 283 gattttaagaataact 299
XX
XX  RESULT 13
XX  AAH94373/C
XX  ID AAH94373 standard; cDNA; 1239 BP.
XX
XX  AAH94373;
XX
XX  05-OCT-2001 (first entry)
XX
XX  Human foetal cDNA, SEQ ID NO: 902.
XX
XX  Human; foetal protein; cytosolic; immunosuppressive; immunostimulant;
XX  neurotropic; neuroprotective; thrombolytic; osteopathic; antiinflammatory;
XX  gene therapy; antisense therapy; cancer; immune disorder;
XX  growth disorder; osteoporosis; thrombolytic disorder;
XX  nervous system disorder; inflammation; expressed sequence tag; EST; ss.
XX
XX  Homo sapiens.
XX
XX  WO200155339-A2.
XX
XX  02-AUG-2001.
XX
XX  25-JAN-2001; 2001WO-US02723.
XX

```

PR 25-JAN-2000; 2000US-0491404.  
PR 15-SEP-2000; 2000US-0663870.  
PR 06-NOV-2000; 2000US-0707351.  
XX  
PA (HISE-) HISEQ INC.  
XX  
PI Yeung G, Ford JE, Boyle BJ, Arterburn MC, Drmanac RA, Tang YT;  
PI Liu C, Asundi V, Zhou P, Werhman T;  
XX  
DR WPI: 2001-465571/50.  
DR P-PSDB: AAM06698.  
XX  
PT Novel fetal proteins useful for the treatment and diagnosis of diseases  
PT associated with dysfunction of the protein e.g. cancers, immune  
PT disorders, growth disorders, thrombolytic disorders, nervous system  
PT disorders and inflammation -  
XX  
PS Claim 1; Page 526-527; 715pp; English.  
XX  
CC The invention relates to novel foetal polypeptides encoded by  
CC polynucleotides comprising one of 477 sequences fully defined in the  
CC specification. The foetal polynucleotides and polypeptides are  
CC useful in the treatment and diagnosis of diseases such as cancers,  
CC immune disorders, growth disorders (e.g. osteoporosis), thrombolytic  
CC disorders, nervous system disorders and inflammation. The present  
CC sequence is a full length cDNA which was assembled using expressed  
CC sequence tags (ESTs) found to be expressed in human foetal tissue  
CC cDNA libraries as seeds.  
XX  
SQ Sequence 1239 BP; 370 A; 187 C; 246 G; 436 T; 0 other;

Query Match 5.4%; Score 17; DB 22; Length 1239;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 34 tgaagaatagaacaat 50  
|||  
DB 1239 TGAAGAAATGACAAAT 1223

RESULT 14  
ABA45382  
ID ABA45382 standard; DNA; 1959 BP.  
XX  
AC ABA45382;  
XX  
DT 01-FEB-2002 (first entry)  
XX  
DE Human breast cell single exon nucleic acid probe #4077.  
XX  
KW Human; microarray; single exon probe; gene expression; breast;  
KW disease; cancer; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157271-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00662.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX

XX  
DR WPI: 2001-496933/54.  
XX  
PT New spatially-addressable set of single exon nucleic acid probes,  
PT useful for measuring gene expression in sample derived from human  
PT breast, comprises number of single exon nucleic acid probes -  
XX  
PS Claim 1; SEQ ID NO 4077; 327pp + sequence listing; English.  
XX  
CC The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human breast and BT 474 cells. The method involves contacting  
CC the probes with a collection of detectably labelled nucleic acids  
CC derived from mRNA of human breast, and then measuring the label  
CC bound to each probe of the microarray. The probes are useful for  
CC verifying the expression of regions of genomic DNA predicted to  
CC encode proteins. They are useful for gene discovery, and for  
CC determining predisposition and/or prognosing breast disease. Gene  
CC expression analysis is useful for assessing the toxicity of chemical  
CC agents on cells. The microarray of this invention presents a far greater  
CC diversity of probes for measuring gene expression, with far less bias  
CC than expressed sequence tag microarrays. The method is suitable for  
CC rapid production of functional information from genomic sequence. The  
CC present sequence is a single exon nucleic acid probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pt\_sequences.  
XX  
SQ Sequence 1959 BP; 521 A; 343 C; 564 G; 531 T; 0 other;

Query Match 5.4%; Score 17; DB 22; Length 1959;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 124 agtgggtgccttgt 140  
|||  
DB 270 agtgggtgccttgt 286

RESULT 15  
ABA55872  
ID ABA55872 standard; DNA; 1959 BP.  
XX  
AC ABA55872;  
XX  
DT 01-FEB-2002 (first entry)  
XX  
DE Human foetal liver single exon nucleic acid probe #4177.  
XX  
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157277-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00669.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI: 2001-483447/52.  
XX

XX Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human fetal liver -  
 XX  
 PS Claim 1: SEQ ID NO 4177; 639pp + sequence listing; English.  
 XX  
 CC The invention relates to a single exon nucleic acid probe for  
 CC measuring human gene expression in a sample derived from human foetal  
 CC liver. The single exon nucleic acid probes may be used for predicting,  
 CC measuring and displaying gene expression in samples derived from human  
 CC fetal liver. The present sequence is a single exon nucleic acid  
 CC probe of the invention.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1959 BP; 521 A; 343 C; 564 G; 531 T; 0 other;

Query Match 5.48; Score 17; DB 22; Length 1959;  
 Best Local Similarity 100.0%; Pred. No. 71;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 124 agtgggtgcttctgt 140  
 |||||  
 Db 270 agtgggtgcttctgt 286

Search completed: September 16, 2002, 06:31:41  
 Job time: 7649 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 06:24:15 ; Search time 90.91 seconds  
(without alignments)  
851.112 Million cell updates/sec

Title: US-09-700-770-4

Perfect score: 315

Sequence: 1 taacactgactcagattt.....tgtgtattacttactatgc 315

Scoring table: OLIGO\_NUC

Searched: 38353 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Issued\_Patents\_NA: \*  
1: /cgn2-6/ptodata/2/1na/5A\_COMB.seq: \*  
2: /cgn2-6/ptodata/2/1na/5B\_COMB.seq: \*  
3: /cgn2-6/ptodata/2/1na/6A\_COMB.seq: \*  
4: /cgn2-6/ptodata/2/1na/6B\_COMB.seq: \*  
5: /cgn2-6/ptodata/2/1na/PCtUS\_COMB.seq: \*  
6: /cgn2-6/ptodata/2/1na/Backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	5.7	9620	4 US-08-952-127-11	Sequence 11, Appl
2	17	5.4	4146	1 US-08-261-822A-15	Sequence 15, Appl
3	17	5.4	4146	5 PCT-US95-07744A-15	Sequence 15, Appl
4	16	5.1	715	4 US-08-991-789A-264	Sequence 264, App
5	16	5.1	715	4 US-09-062-451-264	Sequence 264, App
6	16	5.1	1161	2 US-08-616-857-1	Sequence 1, Appl
7	16	5.1	2831	2 US-09-906-713-1	Sequence 5, Appl
8	16	5.1	6688	4 US-09-381-862-5	Sequence 1, Appl
9	16	5.1	9100	2 US-08-743-637B-27	Sequence 27, Appl
10	16	5.1	9100	3 US-08-526-840B-27	Sequence 27, Appl
11	15	4.8	23	1 US-07-741-940-64	Sequence 64, Appl
12	15	4.8	23	1 US-08-289-348A-64	Sequence 64, Appl
13	15	4.8	23	1 US-08-452-654-64	Sequence 64, Appl
14	15	4.8	23	1 US-08-452-655B-64	Sequence 64, Appl
15	15	4.8	23	3 US-08-450-582-64	Sequence 64, Appl
16	15	4.8	51	2 US-08-350-260A-265	Sequence 265, App
17	15	4.8	54	2 US-08-350-260A-265	Sequence 265, App
18	15	4.8	57	2 US-08-350-260A-267	Sequence 267, App
19	15	4.8	381	2 US-08-844-153-3	Sequence 1, Appl
20	15	4.8	486	1 US-08-828-511-1	Sequence 1, Appl
21	15	4.8	666	3 US-09-081-320-15	Sequence 15, Appl
22	15	4.8	714	4 US-08-896-164-19	Sequence 19, Appl
23	15	4.8	945	1 US-08-253-155A-10	Sequence 10, Appl
24	15	4.8	1006	3 US-08-924-747-15	Sequence 15, Appl
25	15	4.8	1006	4 US-09-247-373B-15	Sequence 15, Appl
26	15	4.8	1006	4 US-09-296-715-15	Sequence 15, Appl
27	15	4.8	1443	2 US-08-962-203-1	Sequence 1, Appl

28	15	4.8	1443	4 US-09-282-125A-1	Sequence 1, Appl
29	15	4.8	1443	4 US-09-273-142-1	Sequence 1, Appl
30	15	4.8	1569	2 US-08-743-637B-176	Sequence 176, App
31	15	4.8	1569	3 US-08-526-840B-176	Sequence 176, App
32	15	4.8	1734	1 US-08-551-437-4	Sequence 4, Appl
33	15	4.8	1734	3 US-09-004-225-4	Sequence 4, Appl
34	15	4.8	1734	4 US-09-084-346-4	Sequence 4, Appl
35	15	4.8	1734	4 US-09-104-704-4	Sequence 4, Appl
36	15	4.8	1886	2 US-08-659-939-1	Sequence 1, Appl
37	15	4.8	1886	2 US-08-850-041-1	Sequence 1, Appl
38	15	4.8	1886	2 US-08-467-963C-5	Sequence 5, Appl
39	15	4.8	1886	2 US-08-838-189D-5	Sequence 5, Appl
40	15	4.8	1886	3 US-08-852-344D-5	Sequence 5, Appl
41	15	4.8	1886	3 US-08-896-500-1	Sequence 1, Appl
42	15	4.8	1886	3 US-08-476-397-1	Sequence 1, Appl
43	15	4.8	1886	3 US-08-973-720-1	Sequence 1, Appl
44	15	4.8	1886	3 US-08-344-639E-5	Sequence 5, Appl
45	15	4.8	1886	3 US-09-262-927-1	Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-08-952-127-11  
; Sequence 11, Application US/08952127  
; Patent No. 6211336  
; GENERAL INFORMATION:  
; APPLICANT: Shiloh, Yosef  
; APPLICANT: Tagle, Danilo A.  
; APPLICANT: Collins, Francis S.  
; TITLE OF INVENTION: ATRAXIA-TELANGIECTASIA GENE  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Kohn & Associates  
; STREET: 30500 No. 6211336thwestern Hwy., Suite 410  
; CITY: Farmington Hills  
; STATE: Michigan  
; COUNTRY: U.S.  
; ZIP: 48334  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/952,127  
; FILING DATE:  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kohn, Kenneth I.  
; REGISTRATION NUMBER: 30,995  
; REFERENCE/DOCKET NUMBER: 2290.00029  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 810-539-5050  
; TELEFAX: 810-539-5055  
; INFORMATION FOR SEQ ID NO. 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 9620 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; ORIGINAL SOURCE:  
; ORGANISM: Mus musculus  
; POSITION IN GENOME:  
; CHROMOSOME/SEGMENT: Chromosome 9, Band 9C  
; US-08-952-127-11  
Query Match 5.7%; Score 18; DB 4; Length 9620;  
Best Local Similarity 100.0%; Pred. NO. 2.4;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 80 tctcttagagatcctacag 97  
|||||  
Db 6071 TCTCTTAGAGATCTACAG 6088

## RESULT 2

US-08-261-822A-15  
; Sequence 15, Application US/08261822A  
; Patent No. 5650553  
; GENERAL INFORMATION:  
; APPLICANT: Ecker, Joseph R. et al.  
; TITLE OF INVENTION: Plant Genes for Sensitivity to Ethylene  
; TITLE OF INVENTION: and Pathogens  
; NUMBER OF SEQUENCES: 82  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5650553tris  
; STREET: One Liberty Place, 46th floor  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/261,822A  
; FILING DATE: 17-JUN-1994  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Beardell, Lori Y.  
; REGISTRATION NUMBER: 34,293  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 568-3439  
; TELEFAX: (215) 568-3439  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4146 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-261-822A-15

Query Match 5.4%; Score 17; DB 1; Length 4146;  
Best Local Similarity 100.0%; Pred No. 7.7;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 176 aaatttgaatatagat 192  
|||||  
Db 1284 AAATTTGAAATAGAT 1300

## RESULT 3

PCT-US95-07744A-15  
; Sequence 15, Application PC/TUS9507744A  
; GENERAL INFORMATION:  
; APPLICANT: Trustees of The University of Pennsylvania  
; TITLE OF INVENTION: Plant Genes for Sensitivity to Ethylene  
; TITLE OF INVENTION: and Pathogens  
; NUMBER OF SEQUENCES: 82  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & Norris  
; STREET: One Liberty Place, 46th floor  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103

## COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/07744A  
; FILING DATE: 15-JUNE-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/261,822  
; FILING DATE: June 17, 1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Beardell, Lori Y.  
; REGISTRATION NUMBER: 34,293  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 568-3439  
; TELEFAX: (215) 568-3439  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4146 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
PCT-US95-07744A-15

Query Match 5.4%; Score 17; DB 5; Length 4146;  
Best Local Similarity 100.0%; Pred. No. 7.7;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 176 aaatttgaatatagat 192  
|||||  
Db 1284 AAATTTGAAATAGAT 1300

## RESULT 4

US-08-991-789A-264/C  
; Sequence 264, Application US/08991789A  
; Patent No. 6225054  
; GENERAL INFORMATION:  
; APPLICANT: Fridakis, Tony N.  
; Smith, John M.  
; Reed, Steven G.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
; TREATMENT AND DIAGNOSIS OF BREAST CANCER  
; NUMBER OF SEQUENCES: 292  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Seed IP Law Group  
; STREET: 701 Fifth Avenue, Suite 6300  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/991,789A  
; FILING DATE: 11-Dec-1997  
; CLASSIFICATION: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Potter, Jane E. R.  
; REGISTRATION NUMBER: 33,332  
; REFERENCE/DOCKET NUMBER: 210121.419C3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031

;; INFORMATION FOR SEQ ID NO: 264:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 715 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; US-08-991-789A-264  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 264:

Query Match 5.1%; Score 16; DB 4; Length 715;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 ttgaaatagattgt 195  
|||||  
Db 289 TTGAAATAGATTGT 274

## RESULT 5

US-09-062-451-264/C  
; Sequence 264, Application US/09062451  
; Patent No. 6344350  
; GENERAL INFORMATION:  
; APPLICANT: Frudakis, Tony N.  
; APPLICANT: Smith, John M.  
; APPLICANT: Reed, Steven G.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
; TITILE OF INVENTION: TREATMENT AND DIAGNOSIS OF BREAST CANCER  
; NUMBER OF SEQUENCES: 297  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentln Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/062,451  
; FILING DATE: 04-APR-1997  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MAKI, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.419C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 264:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 715 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-09-062-451-264

Query Match 5.1%; Score 16; DB 4; Length 715;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 ttgaaatagattgt 195  
|||||  
Db 289 TTGAAATAGATTGT 274

RESULT 6  
US-08-616-857-1/C

;; Sequence 1, Application US/08616857  
;; Patent No. 5905146  
;; GENERAL INFORMATION:  
;; APPLICANT: Lecka-Czerwik, Dr. Beata  
;; TITLE OF INVENTION: NOVEL DNA BINDING PROTEIN S1-3  
;; NUMBER OF SEQUENCES: 7  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Fish & Richardson P.C.  
;; STREET: 225 Franklin Street  
;; CITY: Boston  
;; STATE: MA  
;; COUNTRY: USA  
;; ZIP: 02110-2804  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentln Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/616,857  
;; FILING DATE:  
;; CLASSIFICATION: 514  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Freeman, John F.  
;; REGISTRATION NUMBER: 29,066  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 617/542-5070  
;; TELEFAX: 617/542-8906  
;; TELEX: 200154  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1161 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 2..674  
;; US-08-616-857-1

Query Match 5.1%; Score 16; DB 2; Length 1161;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 ttgaaatagattgt 195  
|||||  
Db 816 TTGAAATAGATTGT 801

RESULT 7  
US-08-906-713-1  
; Sequence 1, Application US/08906713  
; Patent No. 5965704  
; GENERAL INFORMATION:  
; APPLICANT: Lok, Si  
; APPLICANT: Adams, Robyn L.  
; APPLICANT: Jelmeberg, Anna C.  
; APPLICANT: Whitmore, Theodore E.  
; APPLICANT: Farrah, Theresa M.  
; TITLE OF INVENTION: MAMMALIAN ZCYTOR11  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Zymogenetics  
; STREET: 1201 Eastlake Ave East  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98102  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/906,713  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Lunn, Paul G  
REGISTRATION NUMBER: 32,743  
REFERENCE/DOCKET NUMBER: 97-52  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-442-6627  
TELEFAX: 206-442-6678  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2831 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: Coding Sequence  
LOCATION: 34...1755  
OTHER INFORMATION:  
US-08-906-713-1

Query Match 5.1%; Score 16; DB 2; Length 2831;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 44 aacaaatgaatcagt 59  
|||||  
DB 2631 AACAAATGAAATCAGT 2646

RESULT 8  
US-09-381-862-5/c  
Sequence 5, Application US/09381862  
Patent No. 6245906  
GENERAL INFORMATION:  
APPLICANT: Ueyama, Hiroshi  
APPLICANT: Abe, Kanako  
APPLICANT: Keshi, Hiroyuki  
APPLICANT: Matsubisa, Akio  
TITLE OF INVENTION: PROBES FOR THE DIAGNOSIS OF INFECTIONS  
TITLE OF INVENTION: CAUSED BY STREPTOCOCCUS PYOGENES  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive/6300 Sears Tower  
City: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/381,862  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 1997-71077  
FILING DATE: 25-MAR-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP98/01288

FILING DATE: 23-MAR-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Cawley, Jr., Thomas A.  
REGISTRATION NUMBER: 40,944  
REFERENCE/DOCKET NUMBER: 19036/36274  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6688 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus pyogenes  
STRAIN: Clinical Isolate SP-26-46  
US-09-381-862-5

Query Match 5.1%; Score 16; DB 4; Length 6688;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 aaataactttgagaa 39  
|||||  
DB 5369 AAATAACTTTGAGAA 5354

RESULT 9  
US-08-743-637B-27  
Sequence 27, Application US/08743637B  
Patent No. 5994066  
GENERAL INFORMATION:  
APPLICANT: BERGERON, Michel G.  
APPLICANT: PICARD, Francois J.  
APPLICANT: OUELLETTE, Marc  
APPLICANT: ROY, Paul H.  
TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA  
TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND  
TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED  
TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS  
NUMBER OF SEQUENCES: 273  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: QUARLES & BRADY  
STREET: 411 EAST WISCONSIN AVENUE  
City: MILWAUKEE  
STATE: WISCONSIN  
COUNTRY: USA  
ZIP: 53202-4497  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/743,637B  
FILING DATE: 04-NOV-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/526,840  
FILING DATE: 11-SEP-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: BAKER, Jean C.  
REGISTRATION NUMBER: 35,433  
REFERENCE/DOCKET NUMBER: 850586,90012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (414) 277-5000  
TELEFAX: (414)277-5591  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9100 base pairs



```

: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: ORGANISM: Haemophilus influenzae
US-08-743-637B-27

```

```

Query Match      5.1%; Score 16; DB 2; Length 9100;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 234 ggaataacacattcc 249
Db 3581 GGAATACACATTCC 3596

```

```

RESULT 10
US-08-526-840B-27
: Sequence 27, Application US/08526840B
: Patent No. 6001564
: GENERAL INFORMATION:
: APPLICANT: BERGERON, Michel G.
: APPLICANT: OUELLETTE, Marc
: APPLICANT: ROY, Paul H.
: TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
: TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
: TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
: NUMBER OF SEQUENCES: 177
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: OUARLES & BRADY
: STREET: 411 East Wisconsin Avenue
: CITY: Milwaukee
: STATE: Wisconsin
: COUNTRY: USA
: ZIP: 53202-4497
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/526,840B
: FILING DATE: 11-SEP-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/304,732
: FILING DATE: 12-SEP-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: BAKER, Jean C.
: REGISTRATION NUMBER: 35,433
: REFERENCE/DOCKET NUMBER: 850586,90012
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (414) 277-5000
: TELEFAX: (414) 277-5591
: INFORMATION FOR SEQ ID NO: 27:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 9100 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: ORGANISM: Haemophilus influenzae
US-08-526-840B-27

```

```

Query Match      5.1%; Score 16; DB 3; Length 9100;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 234 ggaataacacattcc 249
Db 3581 GGAATACACATTCC 3596

```

```

RESULT 11
US-07-741-940-64
: Sequence 64, Application US/07741940
: Patent No. 5352775
: GENERAL INFORMATION:
: APPLICANT: ALBERTSEN, HANS
: APPLICANT: ANAND, RAKESH
: APPLICANT: CARLSON, MARY
: APPLICANT: GRODEN, JOANNA
: APPLICANT: HEDGE, PHILIP J.
: APPLICANT: JOSLYN, GEOFF
: APPLICANT: KINZLER, KENNETH
: APPLICANT: MARKHAM, ALEXANDER F.
: APPLICANT: NAKAMURA, YUSUKE
: APPLICANT: THLIVERIS, ANDREW
: TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
: TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
: NUMBER OF SEQUENCES: 94
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Banner, Birch, McKie & Beckett
: STREET: 1001 G Street, NW
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20001-4598
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/741,940
: FILING DATE: 19920109
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Kagan, Sarah A.
: REGISTRATION NUMBER: 32,141
: REFERENCE/DOCKET NUMBER: 1107,035574
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-508-9100
: TELEFAX: 202-508-9299
: INFORMATION FOR SEQ ID NO: 64:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 23 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: CDNA
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
US-07-741-940-64

```

```

Query Match      4.8%; Score 15; DB 1; Length 23;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 199 acaataactgagct 213
Db 1 ACAATAACTGAGCT 15

```

```

RESULT 12
US-08-289-548A-64
: Sequence 64, Application US/08289548A
: Patent No. 5648212
: GENERAL INFORMATION:
: APPLICANT: ALBERTSEN, HANS

```

APPLICANT: ANAND, RAKESH  
APPLICANT: CARLSON, MARY  
APPLICANT: GRODEN, JOANNA  
APPLICANT: HEDGE, PHILIP J.  
APPLICANT: JOSLYN, GEOFF  
APPLICANT: KINZLER, KENNETH  
APPLICANT: MARKHAM, ALEXANDER F.  
APPLICANT: NAKAMURA, YUSUKE  
APPLICANT: THLIVERIS, ANDREW  
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
NUMBER OF SEQUENCES: 102  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner & Allegretti, LTD  
STREET: 1001 G Street, NW  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001-4598  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/289,548A  
FILING DATE: 12-AUG-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagau, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 1107,46943  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
INFORMATION FOR SEQ. ID NO: 64:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 23 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
US-08-289-548A-64

Query Match 4.8%; Score 15; DB 1; Length 23;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 199 acaataactgagat 213  
Db 1 ACAATTAAGTGAGT 15

RESULT 13  
US-08-452-654-64  
Sequence 64, Application US/08452654  
Patent No. 5691454  
GENERAL INFORMATION:  
APPLICANT: ALBERTSEN, HANS  
APPLICANT: ANAND, RAKESH  
APPLICANT: CARLSON, MARY  
APPLICANT: GRODEN, JOANNA  
APPLICANT: HEDGE, PHILIP J.  
APPLICANT: JOSLYN, GEOFF  
APPLICANT: KINZLER, KENNETH  
APPLICANT: MARKHAM, ALEXANDER F.  
APPLICANT: NAKAMURA, YUSUKE  
APPLICANT: THLIVERIS, ANDREW  
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
NUMBER OF SEQUENCES: 102  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner & Wilcoff, Ltd.  
STREET: 1001 G Street, NW  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001-4598

NUMBER OF SEQUENCES: 94  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner, Birch, McKie & Beckett  
STREET: 1001 G Street, NW  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001-4598  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/452,654  
FILING DATE: 25-MAY-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/741,940  
FILING DATE: 08-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagau, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 1107,035574  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
INFORMATION FOR SEQ. ID NO: 64:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 23 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
US-08-452-654-64

Query Match 4.8%; Score 15; DB 1; Length 23;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 199 acaataactgagat 213  
Db 1 ACAATTAAGTGAGT 15

RESULT 14  
US-08-452-655B-64  
Sequence 64, Application US/08452655B  
Patent No. 5783666  
GENERAL INFORMATION:  
APPLICANT: ALBERTSEN, HANS  
APPLICANT: ANAND, RAKESH  
APPLICANT: CARLSON, MARY  
APPLICANT: GRODEN, JOANNA  
APPLICANT: HEDGE, PHILIP J.  
APPLICANT: JOSLYN, GEOFF  
APPLICANT: KINZLER, KENNETH  
APPLICANT: MARKHAM, ALEXANDER F.  
APPLICANT: NAKAMURA, YUSUKE  
APPLICANT: THLIVERIS, ANDREW  
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
NUMBER OF SEQUENCES: 102  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner & Wilcoff, Ltd.  
STREET: 1001 G Street, NW  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001-4598

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/452,655B  
FILING DATE: 25-MAY-1995  
CLASSIFICATION: 530  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/289,548  
FILING DATE: 12-AUG-1994  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/741,940  
FILING DATE: 08-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 1107.49964  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
INFORMATION FOR SEQ ID NO: 64:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 23 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
US-08-452-655B-64

Query Match 4.8%; Score 15; DB 1; Length 23;  
Best local Similarity 100.0%; Pred. No. 84;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 199 acaataactgagct 213  
|||||  
Db 1 ACAATAACTGAGCT 15

RESULT 15  
US-08-450-582-64  
Sequence 64, Application US/08450582  
Patent No. 6114124  
GENERAL INFORMATION:  
APPLICANT: ALBERTSEN, HANS  
APPLICANT: ANAND, RAKESH  
APPLICANT: CARLSON, MARY  
APPLICANT: GRODEN, JOANNA  
APPLICANT: HEDGE, PHILIP J.  
APPLICANT: JOSLYN, GEOFF  
APPLICANT: KINZLER, KENNETH  
APPLICANT: MARKHAM, ALEXANDER F.  
APPLICANT: NAKAMURA, YUSUKE  
APPLICANT: THIVIERIS, ANDREW  
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
NUMBER OF SEQUENCES: 102  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner & Witcoff, Ltd.  
STREET: 1001 G Street, NW  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001-4598  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/450,582  
FILING DATE:  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/452,655  
FILING DATE: 25-MAY-1995  
APPLICATION NUMBER: US 08/289,548  
FILING DATE: 12-AUG-1994  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/741,940  
FILING DATE: 08-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 1107.49964  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
INFORMATION FOR SEQ ID NO: 64:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 23 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
US-08-450-582-64

Query Match 4.8%; Score 15; DB 3; Length 23;  
Best local Similarity 100.0%; Pred. No. 84;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 199 acaataactgagct 213  
|||||  
Db 1 ACAATAACTGAGCT 15

Search completed: September 16, 2002, 06:24:21  
Job time: 10634 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 05:18:12 ; Search time 3017.73 Seconds  
(without alignments) 1408.854 Million cell updates/sec

Title: US-09-700-770-4  
Perfect score: 315  
Sequence: 1 taacactgactcagattt.....tggtatatactactatgc 315

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 674847542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

- EST:\*
- 1: em\_estbda:\*
  - 2: em\_estbhm:\*
  - 3: em\_estln:\*
  - 4: em\_estnu:\*
  - 5: em\_estov:\*
  - 6: em\_estpl:\*
  - 7: em\_estro:\*
  - 8: em\_hlc:\*
  - 9: gb\_est1:\*
  - 10: gb\_est2:\*
  - 11: gb\_hlc:\*
  - 12: gb\_gss:\*
  - 13: em\_gss\_hum:\*
  - 14: em\_gss\_inv:\*
  - 15: em\_gss\_pln:\*
  - 16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	264	83.8	563	10	BF335657 RCI-CT046
2	81	25.7	513	10	N49446 yv22a12.t1
3	42	13.3	323	9	AM753759 RCI-CT028
4	21	6.7	573	12	BH11628 RCI-24-2
5	21	6.7	685	12	AO780202 HS-3138.B
6	20	6.3	578	12	A2319774 IM0039F04
7	19	6.0	205	9	AV163511
8	19	6.0	258	10	BG999039
9	19	6.0	300	10	C56866
10	19	6.0	364	10	BM431586
11	19	6.0	376	10	BF711377 MI-P-A1-a
12	19	6.0	416	10	AO669043 HS-5414.A
13	19	6.0	510	12	B1075543 R1075543
14	19	6.0	585	10	BH524918 BCGXT64TF
15	19	6.0	651	9	BB656024
16	19	6.0	680	12	AZ521650 RCI-11-2
17	19	6.0	755	12	BH022291 GH_MBR000

18	19	6.0	788	12	BH250476
19	19	6.0	801	12	BH544046
20	19	6.0	831	12	AZ208098
21	18	5.7	212	12	AZ603387
22	18	5.7	266	9	AV074635
23	18	5.7	288	9	AV072249
24	18	5.7	309	10	B1080659
25	18	5.7	335	9	AA834642
26	18	5.7	364	9	AA206457
27	18	5.7	380	9	AA495343
28	18	5.7	393	10	B1467110
29	18	5.7	393	10	B1790038
30	18	5.7	394	10	B1863517
31	18	5.7	423	10	B1789842
32	18	5.7	436	12	AZ812831
33	18	5.7	450	10	B1441375
34	18	5.7	455	10	BE919766
35	18	5.7	457	10	B1349672
36	18	5.7	457	10	T54305
37	18	5.7	464	10	BF152600
38	18	5.7	471	10	B1043094
39	18	5.7	479	10	BM257055
40	18	5.7	479	12	AO692875
41	18	5.7	486	10	BE677407
42	18	5.7	493	10	BF707366
43	18	5.7	500	9	AU088111
44	18	5.7	500	10	BF440800
45	18	5.7	501	10	BF440803

## ALIGNMENTS

RESULT 1  
BF335657  
LOCUS RCI-CT0469-160800-011-e04 CT0469 Homo sapiens CDNA, mRNA sequence.  
DEFINITION BF335657  
ACCESSION BF335657.1 GI:11306405  
VERSION  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 563)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Britones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

CONTACT: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC1et2-RC1-CT0469-160800-011-e04&t3=2000-08-16&t4=1)

Seq primer: puc 18 forward  
High quality sequence start: 21  
High quality sequence stop: 559.

FEATURES  
source  
1..563  
Location/Qualifiers  
/organism="Homo sapiens"

```

/db_xref="taxon:9606"
/clone_lib="CF0469"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site:1: SmaI; Site:2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 /716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT      184 a      105 c      109 g      165 t
ORIGIN

```

```

Query Match      83.8%; Score 264; DB 10; Length 563;
Best Local Similarity 99.7%; Pred. No. 1.1e-123;
Matches 314; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 taacacgtacacgaattttaagaataaactttgagaataagacaatgaatcagtt 60
    |||||||
DB 180 TAAACACTGACCTCAGATTGTAAGAAATACTTTGAGAAATGAAACAAATGAAATCAGTT 239
    |||||||
QY 61 tctccacacttaagatattctcttagagatctacagctccctccttagggagatacaaa 120
    |||||||
DB 240 TCTCCACCACTTAAGTATGCTCTTGAAGATCTACAGCCTCCCTTTAGGGACATACAAA 299
    |||||||
QY 121 gtcaagtgtgtgcttcttgtagtccacattatataagtagtgatgatacaatt 180
    |||||||
DB 300 GTCAAGTGTGTGCTTGTGTGAGTCCACCTTATATTCAGAGTATGACTACAAAT 359
    |||||||
QY 181 ttgaataatgattgtcacacaataaactgagattataggaacatagtagaagaataa 240
    |||||||
DB 360 TTGAATAATGATTGTTCACCAATTAACCTGAGTTATGAGAAACATCAGTAGAAGAAATA 419
    |||||||
QY 241 caacattccatcccttaagagagatcattcttgcaactgagatgaatttcagtgt 300
    |||||||
DB 420 CAACATTCATCCCTTTACAGAGATCATTTACTTGCACCTCGAATATTTGTCAATGTGT 479
    |||||||
QY 301 attatctactatgc 315
    |||||||
DB 480 ATTATCTACTTATGTC 494
    |||||||

```

```

RESULT 2
N9446 LOCUS      513 bp mRNA linear EST 14-FEB-1996
DEFINITION yv22a12.r1 Soares fetal liver spleen INPLS Homo sapiens cDNA clone
IMAGE:243454 5', mRNA sequence.
ACCESSION N9446
VERSION N9446.1 GI:1190612
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 513)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
,R., Williamson,A., Wohlmann,P. and Wilson,R.
The Mashu-Merck EST Project
Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available royalty-free through LINT; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: 77
High quality sequence stop: 208.
Location/Qualifiers

```

## FEATURES

## source

```

1. 513
/organism="Homo sapiens"
/db_xref="GDB:3792587"
/db_xref="taxon:9606"
/clone_lib="IMAGE:243454"
/clone_lib="Soares fetal liver spleen INPLS"
/sex="male"
/dev_stage="20 week post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
with a modified polylinker; Site:1: Pac I; Site:2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACTGGAAGAATTAATTAAGATCTTTTCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      156 a      102 c      89 g      162 t      4 others
ORIGIN

```

```

Query Match      25.7%; Score 81; DB 10; Length 513;
Best Local Similarity 100.0%; Pred. No. 1.6e-30;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 taacacgtacacgaattttaagaataaactttgagaataagacaatgaatcagtt 60
    |||||||
DB 322 TAAACACTGACCTCAGATTGTAAGAAATACTTTGAGAAATGAAACAAATGAAATCAGTT 381
    |||||||
QY 61 tctccacacttaagatattctcttagagatctacagctccctccttagggagatacaaa 120
    |||||||
DB 382 TCTCCACCACTTAAGTATGCTCTTGAAGATCTACAGCCTCCCTTTAGGGACATACAAA 299
    |||||||

```

```

RESULT 3
AM753759 LOCUS      323 bp mRNA linear EST 28-APR-2000
DEFINITION RC3-CT0283-201199-011-f12 CT0283 Homo sapiens cDNA, mRNA sequence.
ACCESSION AM753759
VERSION AM753759.1 GI:7668691
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 323)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagal,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=RC3&t=RC3-CT0283-
201199-011-f12&t3=1999-11-20&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 23
High quality sequence stop: 262.
Location/Qualifiers

```

## FEATURES

source

1. .323  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="C0283"  
/dev\_stage="Adult"  
/note="Organ: colon; Vector: puc18; Site\_1: Sma1; Site\_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 84 a 59 c 61 g 119 t  
ORIGIN

Query Match 13.3%; Score 42; DB 9; Length 323;  
Best Local Similarity 100.0%; Pred. No. 1e-10;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 18 tttaagaataactttgagaataagacaataaatacagt 59  
|||||  
Db 64 TTTAAGAATTAACCTTTTGAGAAATGACAAATGAATCAGT 23

RESULT 4  
BH11628 573 bp DNA linear GSS 19-JUL-2001  
LOCUS BH11628  
DEFINITION RPCI-24-252C22.TJ RPCI-24 Mus musculus genomic clone RPCI-24-252C22  
ACCESSION BH11628  
VERSION BH11628.1 GI:14946957  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 573)  
Zhao, S., Nieman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M., Tsegaye, G., Geer, K., Krol, M., Shwartsbeyn, A., Gebregorjls, E., Russell, D., de Jong, P. and Fraser, C.M.  
Mouse BAC End Sequences from Library RPCI-24  
Unpublished (1999)  
Other GSSs: RPCI-24-252C22.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
Plate: 252 row: C column: 22  
Seq primer: SP6  
Class: BAC ends.  
Location/Qualifiers  
1. .573  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-24-252C22"  
/clone\_lib="RPCI-24"  
/sex="Male"  
/cell\_type="Spleen/Brain"  
/note="Vector: pTARBAC1; Site\_1: BamH1; Site\_2: BamH1.  
RPCI-24 Mouse BAC library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamH1 sites using KmoI partially digested male C57BL/6J DNA."

FEATURES  
source

BASE COUNT 190 a 103 c 124 g 156 t  
ORIGIN

Query Match 6.7%; Score 21; DB 12; Length 573;  
Best Local Similarity 100.0%; Pred. No. 5.5;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 agatttgaataactttt 34  
|||||  
Db 133 AGATTTAAGAAATTAACCTTT 153

RESULT 5  
AO780202 685 bp DNA linear GSS 02-AUG-1999  
LOCUS AO780202  
DEFINITION HS-3138-B1-H06\_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3138 Col=11 Row=P, DNA sequence.  
ACCESSION AO780202  
VERSION AO780202.1 GI:5683162  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.  
Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
99380589  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones may be purchased from Research Genetics (Info@resgen.com).  
BAC end Web server: <http://www.hsc.washington.edu>  
Plate: 3138 row: P column: 11  
Seq primer: M13 Reverse  
Class: BAC ends  
High quality sequence stop: 685.  
Location/Qualifiers  
1. .685  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate=3138 Col=11 Row=P"  
/clone\_lib="CIT Approved Human Genomic Sperm Library D"  
/sex="male"  
/note="Organ: sperm; Vector: pBelBAC11; BAC clones in E-Coli DH10B"  
E-Coli DH10B"

BASE COUNT 284 a 132 c 94 g 174 t 1 others  
ORIGIN

Query Match 6.7%; Score 21; DB 12; Length 685;  
Best Local Similarity 100.0%; Pred. No. 5.6;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 112 acatacaagaatcagttgtt 132  
|||||  
Db 26 ACATCAAGAAGTCAGTTGTGTT 46

RESULT 6  
AZ319774 578 bp DNA linear GSS 29-SEP-2000  
LOCUS AZ319774  
DEFINITION IM0039F04R Mouse 10kb plasmid UUC1M library Mus musculus genomic clone UUC1M0039F04 R, DNA sequence.

ACCESSION AZ319774  
 VERSION AZ319774.1 GI:10370884  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 578)  
 Dunin, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmood, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A., and Wright, D., Weiss, R.  
 Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
 Unpublished (2000)  
 JOURNAL Contact: Robert B. Weiss  
 COMMENT University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunin@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0039 row: F column: 04  
 Seq primer: CACACAGGAACACGCTATGACC  
 Class: Plasmid ends  
 High quality sequence stop: 578.  
 Location/Qualifiers  
 1..578  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0039F04"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /sex="Male"  
 /lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"  
 /note="vector: PMD42ny; Purified genomic DNA from M. musculus C57BL/6J (male); was obtained from the Jackson Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g14732114/gb1AF129072.1), a copy number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 188 a 106 c 149 g 135 t  
 ORIGIN

Query Match 6.3%; Score 20; DB 12; Length 578;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 tagaagaataacacattc 248  
 ||||||||||||||||  
 DB 63 TAGAAGAAATRCACACATTC 82

RESULT 7  
 LOCUS AV163511  
 DEFINITION AV163511 Mus musculus head C57BL/6J 13-day embryo Mus musculus cDNA

ACCESSION clone 311001B18, mRNA sequence.  
 VERSION AV163511  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 205)  
 Carrinot, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K., Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsuma, H., Oda, H., Owa, C., Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tateo, M., Tomaru, Y., Tomimaga, N., Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y.  
 RIKEN Mouse ESTs  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Chie Owa  
 Genome Science Laboratory  
 RIKEN  
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
 Tel: 81-298-36-9145  
 Fax: 81-298-36-9098  
 Email: genome-res@rtc.riken.go.jp  
 Thermostabilization and thermostabilization of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))  
 Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))  
 Please visit our web site (http://genome.rtc.riken.go.jp) for further details.  
 Location/Qualifiers  
 1..205  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="311001B18"  
 /clone\_lib="Mus musculus head C57BL/6J 13-day embryo"  
 /sex="mixed"  
 /tissue\_type="head"  
 /dev\_stage="13-day embryo"  
 /base\_count="31 c 37 g 73 t"

BASE COUNT 64 a 31 c 37 g 73 t  
 ORIGIN

Query Match 6.0%; Score 19; DB 9; Length 205;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 ttctccacaccttaagta 77  
 ||||||||||||||||  
 DB 17 TTTCTCCACACACTTAAGTA 35

RESULT 8  
 LOCUS BG999039/c  
 DEFINITION PMO-HR0913-250401-015-C04 HT0913 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BG999039  
 VERSION BG999039.1 GI:14403111  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 258)  
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jorgensen, C.V., O'Hare, M.J., Soares, P., Brentani, R.R., Reis, L.F., de Souza, S.J. and



TITLE  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

JOURNAL  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

COMMENT  
2020263  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1=PM0&t2=PM0-HT0913-  
250401-015-C04&t3=2001-04-25&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 22  
High quality sequence stop: 258.

FEATURES  
source  
location/Qualifiers  
1..258  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="HT0913"  
/dev\_stage="Adult"  
/note="Organ: head\_neck; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

BASE COUNT  
85 a 53 c 55 g 65 t

ORIGIN

Query Match  
Best Local Similarity 100.0%; Score 19; DB 10; Length 258;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 13 cagatttaagaataact 31  
|||||

Db 226 CAGATTTAAGAAATTA 208

RESULT 9  
C56866 300 bp mRNA linear EST 22-SEP-1997  
LOCUS  
DEFINITION C56866 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA  
clone yk278b6 3', mRNA sequence.  
ACCESSION C56866  
VERSION C56866.1 GI:2414897  
KEYWORDS EST.  
SOURCE  
ORGANISM Caenorhabditis elegans.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides  
; Rhabditidae; Pelodirinae; Caenorhabditis.  
1 (bases 1 to 300)  
Kohara, Y., Motonashi, T., Tabara, H., Watanabe, H., Sugimoto, A., Sano  
, M., Miyata, A. and Nishigaki, A.  
Expression map of the C.elegans genome  
Unpublished (1996)  
Contact: Yuji Kohara  
Genome Biology Lab.  
National Institute of Genetics  
Yata 1111, Mishima, Shizuoka 411, Japan  
Tel: 81-559-81-6854  
Fax: 81-559-81-6855  
Email: ykohara@lab.nig.ac.jp.  
Location/Qualifiers

FEATURES  
source  
1..300  
/organism="Caenorhabditis elegans"

/strain="CB1489 him-8(e1489)"  
/db\_xref="taxon:6239"  
/clone\_lib="yk278b6"  
/clone\_lib="Yuji Kohara unpublished cDNA"  
/sex="hermaphrodite, male"  
/tissue\_type="whole animal"  
/dev\_stage="varied"

BASE COUNT  
138 a 29 c 59 g 70 t 4 others

ORIGIN

Query Match  
Best Local Similarity 100.0%; Score 19; DB 10; Length 300;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 168 atgactacaatttga 186  
|||||

Db 71 ATGACTACAAATTTGAAA 89

RESULT 10  
BM431586 364 bp mRNA linear EST 31-JAN-2002  
LOCUS  
DEFINITION 1duo24d11.ab1 Bos taurus Duodenum #1 library Bos taurus cDNA, mRNA  
sequence.  
ACCESSION BM431586  
VERSION BM431586.1 GI:18453308  
KEYWORDS EST.  
SOURCE  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
1 (bases 1 to 364)  
Hansen, C., Fu, A., Weng, Y., Li, C., Okine, E., Sensen, C.W., Gordon  
, P.M.K. and Moore, S.S.  
Gene Expression Profiling of the Bovine Gastrointestinal Tract  
Unpublished (2002)  
Contact: Dr. Stephen Moore  
Beef Genomics Laboratory  
Dept of AFNS, University of Alberta  
410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada  
Tel: 780 492 0169  
Fax: 780 492 4265  
Email: smoores@afns.ualberta.ca  
Insert Length: 364 Std Error: 0.00  
POLYA=Yes.

FEATURES  
source  
location/Qualifiers  
1..364  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="Bos taurus Duodenum #1 library"  
/tissue\_type="Smooth muscle"  
/cell\_type="Simple columnar epithelial"  
/dev\_stage="Young adult"  
/lab\_host="X11-BlueMF's strain"  
/note="Organ: Intestine/Duodenum; Vector: Uni-ZAPvXR;  
Site\_1: EcoRI; Site\_2: Xho I"

BASE COUNT  
123 a 69 c 62 g 110 t

ORIGIN

Query Match  
Best Local Similarity 100.0%; Score 19; DB 10; Length 364;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 169 tgaactacaatttga 187  
|||||

Db 343 TGACTACAAATTTGAAA 361

RESULT 11  
BF711377/c

LOCUS BF711377 376 bp mRNA linear EST 02-JAN-2001  
 DEFINITION MI-P-AI-acm-g-11-1-UM.s1 MI-P-AI Sus scrofa cDNA clone  
 ACCESSION MI-P-AI-acm-g-11-1-UM 3', mRNA sequence.  
 VERSION BF711377  
 KEYWORDS EST.  
 SOURCE pig.  
 ORGANISM Sus scrofa  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.  
 REFERENCE 1 (bases 1 to 376)  
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 COMMENT Contact: Tugayle CK  
 Molecular Genetics Laboratory, Department of Animal Science  
 Iowa State University  
 201 Kildee Hall, Ames, IA 50011-3150, USA  
 Tel: 5152944252  
 Fax: 5152942401  
 Email: ctugayle@iastate.edu  
 The sequence contained an oligo-dT track that was present in the  
 oligonucleotide that was used to prime the synthesis of first  
 strand cDNA and therefore this may represent a bonafide poly A  
 tail. The sequence tag present in the cDNA between the NotI site  
 and the oligo-dT track served to verify it as a clone from the  
 non-normalized anterior pituitary at estrus day 0 library cDNA  
 library preparation: RJ Woods, JA Green, RS Prather, S142 Animal  
 Science Research Center, Department of Animal Science, University  
 of Missouri-Columbia, 65211 Clone distribution: clones will be  
 available through Research Genetics (www.resgen.com)  
 Seq primer: M13 Forward  
 POLYA=Yes.

FEATURES  
 source Location/Qualifiers  
 1..376  
 /organism="Sus scrofa"  
 /strain="crossbred"  
 /db\_xref="taxon:9823"  
 /clone="MI-P-AI-acm-g-11-1-UM"  
 /clone\_lib="MI-P-AI"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: pRT73D-Pac (Pharmacia) with a modified  
 polylinker. Site\_1: Not I; Site\_2: EcoRI; The MI-P-AI  
 library is derived from anterior pituitary at estrus day  
 0. For a detailed description of the library from which  
 this clone was derived, please visit our web site at  
 http://pigest.genome.iastate.edu/  
 TAG\_LIB=MI-P-AI  
 TAG\_TISSUE=anterior pituitary at estrus day 0  
 TAG\_SEQ=TAAGGC  
 TAG\_SEO=TAAGGC"

BASE COUNT 133 a 65 c 57 g 121 t  
 ORIGIN

Query Match 6.0%; Score 19; DB 10; Length 376;  
 Best Local Similarity 100.0%; Pred. No. 54;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 162 gtatgtatgactacaatt 180  
 ||||||||||||||||  
 Db 280 GTAGGTATGACTACAAATT 262

RESULT 12 416 bp DNA linear GSS 24-JUN-1999  
 LOCUS AO669043/c  
 DEFINITION HS\_5414\_A2\_A09\_SP6E RPCI-11 Human Male BAC Library Homo sapiens  
 genomic clone Plate=990 Col=18 Row=A, DNA sequence.  
 ACCESSION AO669043  
 VERSION AO669043.1 GI:5201789  
 KEYWORDS GSS.

SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 416)  
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
 Hood,L.  
 TITLE Sequence-tagged connectors: A sequence approach to mapping and  
 scanning the human genome  
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
 MEDLINE 99380589  
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu  
 Clones are derived from the human BAC library RPCI-11. For BAC  
 library availability, please contact Pieter de Jong  
 (pieterdejong.med.buifalo.edu). Clones may be purchased from  
 BACPAC Resources (http://bacpac.med.buifalo.edu/ordering.bac.htm)  
 or from Research Genetics (info@resgen.com). BAC end Web Server:  
 http://www.htsc.washington.edu  
 Plate: 990 row: A column: 18  
 Seq primer: SP6  
 Class: BAC ends  
 High quality sequence stop: 416.

FEATURES  
 source Location/Qualifiers  
 1..416  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="Plate=990 Col=18 Row=A"  
 /clone\_lib="RPCI-11 Human Male BAC Library"  
 /sex="male"  
 /note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
 Male blood DNA was isolated from one randomly chosen donor  
 and partially digested with a combination of EcoRI and  
 EcoRI Methylase. Size selected DNA was cloned into the  
 pBACe3.6 vector at EcoRI sites"

BASE COUNT 187 a 64 c 64 g 101 t  
 ORIGIN

Query Match 6.0%; Score 19; DB 12; Length 416;  
 Best Local Similarity 100.0%; Pred. No. 55;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 280 tcagataattgtcatgt 298  
 ||||||||||||||||  
 Db 125 TCAGTATATTGTCTATGT 107

RESULT 13 510 bp mRNA linear EST 11-DEC-2001  
 LOCUS BJ075543/c  
 DEFINITION BJ075543 NIBB Mochli normalized Xenopus tailbud library Xenopus  
 laevis cDNA clone XL057004 5', mRNA sequence.  
 ACCESSION BJ075543  
 VERSION BJ075543.1 GI:17520459  
 KEYWORDS EST.  
 SOURCE African clawed frog.  
 ORGANISM Xenopus laevis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 Xenopodinae; Xenopus.  
 REFERENCE 1 (bases 1 to 510)  
 AUTHORS Kitayama,A., Terasaka,C., Mochli,M., Ueno,N., Shin-I,T. and Kohara  
 Y.  
 TITLE Expressed genes in X. laevis embryo  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Tadasu Shin-I

Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp.  
Location/Qualifiers

FEATURES  
source  
1. .510  
/organism="Xenopus laevis"  
/db\_xref="taxon:8355"  
/clone\_id="NIBB Mochi normalized Xenopus tailbud library"  
/tissue\_type="whole embryo"  
/dev\_stage="stage 25"

BASE COUNT 156 a 107 c 110 g 137 t  
ORIGIN

Query Match 6.0%; Score 19; DB 10; Length 510;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 tagaagaataacacatt 247  
|||||  
Db 355 TAGAAGAAATACACATT 337

RESULT 14  
BH524918 585 bp DNA linear GSS 13-DEC-2001  
LOCUS BOGXT64TF BOG Brassica oleracea genomic clone BOGXT64, DNA  
DEFINITION  
sequence.  
BH524918  
ACCESSION BH524918.1 GI:17733003  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

Brassica oleracea.  
Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
1 (bases 1 to 585)  
Town,C.D., Van Aken,S., Uterback,T. and Fraser,C.M.  
Whole genome shotgun sequencing of Brassica oleracea  
Unpublished (2001)  
Other GSSs: BOGXT64TR

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Contact: Chris Town  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TF  
Class: sheared ends.

FEATURES  
source  
1. .585  
/organism="Brassica oleracea"  
/strain="T01000DH3"  
/db\_xref="taxon:3712"  
/clone\_id="BOGXT64"  
/clone\_lib="BOGX"  
/note="Vector: PHOSI, Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOSI using BstXI linkers"  
BASE COUNT 168 a 116 c 110 g 171 t  
ORIGIN

Query Match 6.0%; Score 19; DB 12; Length 585;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 50 tgaatcagttctccacc 68

Db 490 TGAATCAGTTCTCCACC 472  
|||||

RESULT 15  
BB656024 651 bp mRNA linear EST 26-OCT-2001  
LOCUS BB656024 RIKEN full-length enriched, 12 days embryo spinal ganglion  
DEFINITION Mus musculus cDNA clone D130056E14 5', mRNA sequence.  
BB656024  
VERSION BB656024.1 GI:16489852  
KEYWORDS  
SOURCE  
ORGANISM

house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 651)  
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,  
Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,  
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,  
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,  
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,  
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,  
Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)  
Unpublished (2001)

CONTACT: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gscc.riken.go.jp  
URL: http://genome-gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,  
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)

wagii,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,  
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and  
Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,  
Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamana,K., Iizawa,  
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and  
Hayashizaki,Y.  
Computational Analysis of Full-length Mouse cDNAs Compared with  
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)  
Please visit our web site (http://genome-gsc.riken.go.jp) for  
further details.  
e mouse tissues.

FEATURES  
source  
1. .651  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone\_id="D130056E14"  
/clone\_lib="RIKEN full-length enriched, 12 days embryo  
spinal ganglion"  
/tissue\_type="spinal ganglion"  
/dev\_stage="12 days embryo"  
/lab\_host="DH10B"  
/note="Site\_1: SalI; Site\_2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5',  
GAGAGAGAGAGAGCGGCCGCACTCGAGTTTCTTTTCTTTTCTTTVN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. Second strand cDNA was prepared with the  
primer adapter of sequence [5',  
GAGAGAGAGATTCTCGAGTTTAAATTAATCCCGCCCCCGCC 3']. cDNA  
was cleaved with BamH1 and XhoI. Vector: a modified  
pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 217 a. 103 c 111 g 220 t  
ORIGIN

Query Match 6.0%; Score 19; DB 9; Length 651;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 197 acacaataaactgagatt 215  
|||||  
DB 146 ACACATAAATACTGGAGTTT 164

Search completed: September 16, 2002, 05:18:16  
Job time: 8389 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 03:26:41 : Search time 89 Seconds  
(without alignments)  
869.377 Million cell updates/sec

Title: US-09-700-770-4

Perfect score: 315  
Sequence: 1 taacactgcctcagattt.....tgygtattacttactatgc 315

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/lna/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/lna/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/lna/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/lna/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/lna/PTC05.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/lna/Backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31.2	9.9	9919	US-08-880-179-1	Sequence 1, Appl
2	31	9.8	747	US-08-680-726A-87	Sequence 87, Appl
3	31	9.8	747	US-09-092-409-87	Sequence 87, Appl
4	31	9.8	749	US-08-680-726A-85	Sequence 85, Appl
5	31	9.8	749	US-08-680-726A-86	Sequence 86, Appl
6	31	9.8	749	US-09-092-409-85	Sequence 85, Appl
7	31	9.7	749	US-09-092-409-86	Sequence 86, Appl
8	30.6	9.7	6082	US-09-439-313-535	Sequence 535, App
9	30.2	9.6	386	US-08-906-769-142	Sequence 142, App
10	30.2	9.6	386	US-08-906-616-142	Sequence 142, App
11	30.2	9.6	386	US-08-639-075A-142	Sequence 142, App
12	30.2	9.6	386	US-09-012-431-142	Sequence 142, App
13	30.2	9.6	386	US-09-012-692-142	Sequence 142, App
14	30.2	9.6	386	US-08-906-613-142	Sequence 142, App
15	30	9.5	3891	US-08-480-604A-27	Sequence 27, Appl
16	30	9.5	3891	US-08-405-496A-27	Sequence 27, Appl
17	30	9.5	3891	US-08-915-136-27	Sequence 27, Appl
18	29.8	9.5	6140	US-09-439-313-536	Sequence 536, App
19	29.6	9.4	665	US-08-998-416-937	Sequence 937, App
20	29.6	9.4	10207	US-08-920-812-2	Sequence 2, Appl
21	29.6	9.4	10207	US-08-920-827-2	Sequence 2, Appl
22	29.6	9.4	10207	US-08-921-177-2	Sequence 2, Appl
23	29.6	9.4	10207	US-08-362-577C-2	Sequence 2, Appl
24	29.6	9.4	10207	US-08-920-828-2	Sequence 2, Appl
25	29.4	9.3	51952	US-08-947-823-1	Sequence 1, Appl
26	29.2	9.3	758	US-08-817-913-10	Sequence 10, Appl
27	29.2	9.3	855	US-08-817-913-11	Sequence 11, Appl

28	29.2	9.3	859	US-08-817-913-12	Sequence 12, Appl
29	29.2	9.3	1214	US-08-817-913-13	Sequence 13, Appl
30	29.2	9.3	1232	US-08-817-913-14	Sequence 14, Appl
31	29.2	9.3	1352	US-08-817-913-15	Sequence 15, Appl
32	29.2	9.3	1734	US-08-817-913-16	Sequence 16, Appl
33	29.2	9.3	1920	US-08-817-913-17	Sequence 17, Appl
34	28.6	9.1	797	US-08-752-133-1	Sequence 1, Appl
35	28.6	9.1	4307	US-08-944-594-1	Sequence 1, Appl
36	28.6	9.1	6216	US-09-213-053-1	Sequence 1, Appl
37	28.6	9.1	14655	US-08-687-080-59	Sequence 59, Appl
38	28.4	9.0	1789	US-08-895-521-2	Sequence 2, Appl
39	28.4	9.0	1789	US-09-235-218-2	Sequence 2, Appl
40	28.4	9.0	1981	US-09-010-398-3	Sequence 3, Appl
41	28.4	9.0	1981	US-09-366-260-3	Sequence 3, Appl
42	28.4	9.0	8224	US-09-010-398-14	Sequence 14, Appl
43	28.4	9.0	8224	US-09-366-260-14	Sequence 14, Appl
44	28.2	9.0	72604	US-09-268-992-7	Sequence 7, Appl
45	28	8.9	1529	US-08-448-110-1	Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-08-880-179-1/C  
; Sequence 1, Application US/08880179  
; Patent No. 6091004  
; GENERAL INFORMATION:  
; APPLICANT: Ryals, John  
; APPLICANT: Delaney, Terry  
; APPLICANT: Friedlich, Leslie  
; APPLICANT: Weymann, Kristlanna  
; APPLICANT: Layton, Kay  
; APPLICANT: Ellis, Daniel  
; APPLICANT: Uknes, Scott  
; APPLICANT: Jesse, Taco  
; APPLICANT: Vos, Pieter  
; TITLE OF INVENTION: GENE ENCODING A PROTEIN INVOLVED IN THE  
; TITLE OF INVENTION: SIGNAL TRANSDUCTION CASCADE LEADING TO SYSTEMIC ACQUIRED RE  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: No. 6091004artis Corporation  
; STREET: 520 White Plains Road, P.O. Box 2005  
; CITY: Tarrytown  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10591  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/880,179  
; FILING DATE:  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meigs, J. Timothy  
; REGISTRATION NUMBER: 38,241  
; REFERENCE/DOCKET NUMBER: CGC 1909  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (919) 541-8587  
; TELEFAX: (919) 541-8689  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 9919 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO

US-08-880-179-1

Query Match 9.9%; Score 31.2; DB 3; Length 9919;  
Best Local Similarity 46.1%; Pred. No. 6.5;  
Matches 105; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 72 taagataatcttagagatctacagccctttagggagacatacaagtcagttgt 131  
DB 7556 TAAAAAATATTAGTGTAGTGTGAACATTTAAACAAAGATATACTAGAGTTT 7497  
QY 132 tgccttggtgagccacccttataatcaagtagtgatgactacaatttgaataa 191  
DB 7496 TTCCTCTTTAATTCATATATATCTTAAATTTAAATTAAGAACTAAGAAA 7437  
QY 192 ttgtcacacaataaactgagttatggaacatcagtagaaggaataacacattcat 251  
DB 7436 GTACTAGATAGTAAAGAAATTAAGAAAGATTAATCTGTGAAAAATGGAATTTACGT 7377  
QY 252 cccttacaagatcattacttgcacatcagataattgtcatgtg 299  
DB 7376 AAGTTAATTAGAAATTAATCTGCTAAACATGATTAATCTTTTGTG 7329

RESULT 2

US-08-680-726A-87  
Sequence 87, Application US/08680726A  
Patent No. 5804197  
GENERAL INFORMATION:  
APPLICANT: Haanes, Elizabeth J.  
TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES  
NUMBER OF SEQUENCES: 92  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/680,726A  
FILING DATE: 12-JUL-1996  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-46-C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ. ID NO: 87:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 747 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..747  
US-08-680-726A-87

Query Match 9.8%; Score 31; DB 1; Length 747;  
Best Local Similarity 51.9%; Pred. No. 3.3;  
Matches 70; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 174 acaatttgaataatagatgtgcacacataaactgagttatggaacatcagtagaa 233  
DB 20 AAAATTAATTATGAGCGCTGTTATAGCTTAATCTTATATATGGAACCAATCTCTA 79  
QY 234 ggaataacaacatccatccctttagacagagatcattactgaaccagagataattgt 293  
DB 80 CAATAATTAATACCTTCTTGATATGATGATCAGCATAGCTTAATATCATGATGATTAA 139  
QY 294 catgtatattactca 308  
DB 140 TAACATATATTATCTA 154

RESULT 3

US-09-092-409-87  
Sequence 87, Application US/09092409  
Patent No. 6158478  
GENERAL INFORMATION:  
APPLICANT: Haanes, Elizabeth J.  
TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES  
NUMBER OF SEQUENCES: 92  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/092,409  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/680,726  
FILING DATE: 12-JUL-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-46-C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ. ID NO: 87:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 747 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..747  
US-09-092-409-87

Query Match 9.8%; Score 31; DB 3; Length 747;  
Best Local Similarity 51.9%; Pred. No. 3.3;  
Matches 70; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 174 acaatttgaataatagatgtgcacacataaactgagttatggaacatcagtagaa 233  
DB 20 AAAATTAATTATGAGCGCTGTTATAGCTTAATCTTATATATGGAACCAATCTCTA 79  
QY 234 ggaataacaacatccatccctttagacagagatcattactgaaccagagataattgt 293  
DB 80 CAATAATTAATACCTTCTTGATATGATGATCAGCATAGCTTAATATCATGATGATTAA 139



```

; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 86:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/680,726
; FILING DATE: 12-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-46-C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 85:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 749 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-092-409-85

Query Match          9.8%; Score 31; DB 3; Length 749;
Best Local Similarity 51.9%; Pred. No. 3.3;
Matches 70; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 174 acaatttgaataagatgtcacacaataaactgagttttagaacaatcagtagaa 233
DB 730 AAATATTATTGAGCGCTGTTATGCTTAATCTTTATTATATGGAACCAATCTCTA 671
QY 234 ggaatatacaatccatcccttaccagagatcatctactgaactcaggataattgt 293
DB 670 CAAATATTAACTTCTTGTGATTGATGATCAGCATACGTTAATATTCATGATGATTAA 611
QY 294 catgtgtattatctcta 308
DB 610 TAACATATTATATCTA 596

RESULT 7
; Sequence 86, Application US/09092409
; Patent No. 6159478
; GENERAL INFORMATION:
; APPLICANT: Haanes, Elizabeth J.
; APPLICANT: Frank, Rexam S.
; TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES
; NUMBER OF SEQUENCES: 92
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/092,409
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/680,726
; FILING DATE: 12-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-46-C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
```

```

; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 749 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-092-409-86

Query Match          9.8%; Score 31; DB 3; Length 749;
Best Local Similarity 51.9%; Pred. No. 3.3;
Matches 70; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 174 acaatttgaataagatgtcacacaataaactgagttttagaacaatcagtagaa 233
DB 20 AAATATTATTGAGCGCTGTTATGCTTAATCTTTATTATATGGAACCAATCTCTA 79
QY 234 ggaatatacaatccatcccttaccagagatcatctactgaactcaggataattgt 293
DB 80 CAAATATTAACTTCTTGTGATTGATGATCAGCATACGTTAATATTCATGATGATTAA 139
QY 294 catgtgtattatctcta 308
DB 140 TAACATATTATATCTA 154

RESULT 8
; US-09-439-313-535/c
; Sequence 535, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yugu
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 535
; LENGTH: 6082
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-439-313-535

Query Match          9.7%; Score 30.6; DB 4; Length 6082;
Best Local Similarity 68.9%; Pred. No. 8.3;
Matches 42; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 3 aacacgactcagattttagaataacttlttagaataagacaatgaatcagtttc 62
DB 4387 ACCTGATTTGATTTTAAACCAACTTGTGACATTTAAATAAATGAACCATTTTG 4328
QY 63 t 63
DB 4327 T 4327

RESULT 9
; US-08-906-769-142/c
```



```
; Sequence 142, Application US/08906769
; Patent No. 6077687
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Silver, Gary
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
; TITLE OF INVENTION: MOLECULES AND USES THEREOF
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/906,769
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/639, 075
; FILING DATE: 24-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 142:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 386 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..384
; OTHER INFORMATION:
; OTHER INFORMATION: R."
US-08-906-769-142

Query Match          9.6%; Score 30.2; DB 3; Length 386;
Best Local Similarity 49.7%; Pred. No. 4.6;
Matches 77; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
```

```
; Sequence 142, Application US/08906616
; Patent No. 6121035
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Silver, Gary
; TITLE OF INVENTION: FLEA AMINOPEPTIDASE PROTEINS AND USES THEREOF
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/906,616
; FILING DATE: 05-AUG-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C2-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 142:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 386 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..384
; OTHER INFORMATION:
; OTHER INFORMATION: R."
US-08-906-616-142

Query Match          9.6%; Score 30.2; DB 3; Length 386;
Best Local Similarity 49.7%; Pred. No. 4.6;
Matches 77; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
```

```
Oy 139 gtgagtcaccattatccaagtagtatacacaatttgaataatagatgtcac 198
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 280 GTACATTTACAGCCAAATAATATAGTATGAGAAACCAATCCACTTACCC 221
Oy 199 acaataaactgagattatgtaaacatcagtagaagaatacaatccatccctta 258
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 220 AACCTGCCACGGTCTGCTATGAAACCAATGCGTTTCATTCCTTCATCAATTA 161
Oy 259 cagagatcattactgcaactcagataattgt 293
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 160 CAGGTTGATAGATTGAGTTAAACAAATTTT 126
```

RESULT 10  
US-08-906-616-142/c

```
Oy 139 gtgagtcaccattatccaagtagtatacacaatttgaataatagatgtcac 198
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 280 GTACATTTACAGCCAAATAATATAGTATGAGAAACCAATCCACTTACCC 221
Oy 199 acaataaactgagattatgtaaacatcagtagaagaatacaatccatccctta 258
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 220 AACCTGCCACGGTCTGCTATGAAACCAATGCGTTTCATTCCTTCATCAATTA 161
Oy 259 cagagatcattactgcaactcagataattgt 293
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 160 CAGGTTGATAGATTGAGTTAAACAAATTTT 126
```

RESULT 11  
US-08-639-075A-142/c  
; Sequence 142, Application US/08639075A  
; Patent No. 6150125  
; GENERAL INFORMATION:  
; APPLICANT: Grieve, Robert B.

```

;
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Silver, Gary
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
; TITLE OF INVENTION: MOLECULES AND USES THEREOF
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/639,075A
; FILING DATE: 24-APR-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 142:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 386 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..384
; OTHER INFORMATION: /note="At pos. bp 3, change A to
; OTHER INFORMATION: R."
; US-08-639-075A-142

```

```

Query Match          9.6%; Score 30.2; DB 3; Length 386;
Best Local Similarity 49.7%; Pred. No. 4.6;
Matches 77; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
QY 139 gttgagtcaccattatattcaagtagtatacacaattttgaataatgtagtcac 198
    || || || || || || || || || || || || || || || || || || || ||
DB 280 GTACATTTACACCAATAAATATGATGTACGGAACCAATCCACTACTTATACCCC 221
QY 199 acaataaactgagttatagaacatcagtagaagaaatacaatccatccctta 258
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 220 AACCTGCCACGGTTGCTATTTGAACCAATTAGCCGTTTCGATTCCTTCATTAAGCTT 161
QY 259 cagagatcattacttgcaactcagagataattgt 293
    || | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 160 CAGGTTTGATGATTTTGAGTTAAACCAATTTT 126

```

```

RESULT 12
US-09-012-431-142/c
; Sequence 142, Application US/09012431
; Patent No. 6180383
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.

```

```

;
; Stiegler, Gary
; Silver, Gary
; Silver, Patrick J.
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
; TITLE OF INVENTION: MOLECULES AND USES THEREOF
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/012,431
; FILING DATE: 23-Jan-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/639,075
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 142:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 386 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..384
; OTHER INFORMATION: /note="At pos. bp 3, change A to
; OTHER INFORMATION: R."
; SEQUENCE DESCRIPTION: SEQ ID NO: 142:
; US-09-012-431-142

```

```

Query Match          9.6%; Score 30.2; DB 4; Length 386;
Best Local Similarity 49.7%; Pred. No. 4.6;
Matches 77; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
QY 139 gttggtccaccattatattcaagtagtatacacaattttgaataatgtagtcac 198
    || || || || || || || || || || || || || || || || || || || ||
DB 280 GTACATTTACACCAATAAATATGATGTACGGAACCAATCCACTACTTATACCCC 221
QY 199 acaataaactgagttatagaacatcagtagaagaaatacaatccatccctta 258
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 220 AACCTGCCACGGTTGCTATTTGAACCAATTAGCCGTTTCGATTCCTTCATTAAGCTT 161
QY 259 cagagatcattacttgcaactcagagataattgt 293
    || | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 160 CAGGTTTGATGATTTTGAGTTAAACCAATTTT 126

```

```

RESULT 13
US-09-012-692-142/c
; Sequence 142, Application US/09012692
; Patent No. 6214579
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley

```

```

APPLICANT: Frank, Glenn R.
APPLICANT: Stiegler, Gary
APPLICANT: Gaines, Patrick J.
APPLICANT: Silver, Gary
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
TITLE OF INVENTION: MOLECULES AND USES THEREOF
NUMBER OF SEQUENCES: 190
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/906,613
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/639,075
FILING DATE: 24-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 142:
SEQUENCE CHARACTERISTICS:
LENGTH: 386 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..384
OTHER INFORMATION:
OTHER INFORMATION: R."
US-08-906-613-142

Query Match 9.6%; Score 30.2; DB 4; Length 386;
Best Local Similarity 49.7%; Pred. No. 4.6;
Matches 77; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 139 gtgagtcctccacctattctcaagtagagtaagacacaaatttgaataatgattgtcac 198
   || || || || || || || || || || || || || || || || || || || || ||
Db 280 GTACTTTACAGCCATTAATTAATTAGGTATGACGAAACCAATCCACTTATTAACCC 221

QY 199 acaataaacctgaggttatgtgaacatcagtagaagaaataacaacattccatccotta 258
   || || || || || || || || || || || || || || || || || || || || ||
Db 220 AACCGCCACAGCGTTCTGTTAGACACATTAGCCGTTTCGATTCCTTCATCAATTAAGCTT 161

QY 259 cagagatcatctacttgcaactcaggaataatttgt 293
   || || || || || || || || || || || || || || || || || || || || ||
Db 160 CAGGTTGATGATTTTGAGTTAAACCAATTTT 126

RESULT 15
US-08-480-604A-27/c
; Sequence 27, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.

```

```

1  APPLICANT:  FIRCA, JOSEPH R.
2  APPLICANT:  STAFFORD, DOUGLAS C.
3  TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
4  TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
5  NUMBER OF SEQUENCES:  32
6  CORRESPONDENCE ADDRESSES:
7  ADDRESSEE:  MEDLEN & CARROLL, LLP
8  STREET:  220 MONTGOMERY STREET, SUITE 2200
9  CITY:  SAN FRANCISCO
10 STATE:  CALIFORNIA
11 COUNTRY:  UNITED STATES OF AMERICA
12 ZIP:  94104
13
14 COMPUTER READABLE FORM:
15 MEDIUM TYPE:  Floppy disk
16 COMPUTER:  IBM PC compatible
17 OPERATING SYSTEM:  PC-DOS/MS-DOS
18 SOFTWARE:  PatentIn Release #1.0, Version #1.30
19 CURRENT APPLICATION DATA:
20 APPLICATION NUMBER:  US/08/480,604A
21 FILING DATE:  07-JUN-1995
22 CLASSIFICATION:  424
23
24 PRIOR APPLICATION DATA:
25 APPLICATION NUMBER:  US 08/422,711
26 FILING DATE:  14-APR-1995
27
28 PRIOR APPLICATION DATA:
29 APPLICATION NUMBER:  US 08/405,496
30 FILING DATE:  16-MAR-1995
31
32 PRIOR APPLICATION DATA:
33 APPLICATION NUMBER:  US 08/329,154
34 FILING DATE:  25-OCT-1994
35
36 PRIOR APPLICATION DATA:
37 APPLICATION NUMBER:  US 08/161,907
38 FILING DATE:  02-DEC-1993
39
40 PRIOR APPLICATION DATA:
41 APPLICATION NUMBER:  US 07/985,321
42 FILING DATE:  04-DEC-1992
43
44 PRIOR APPLICATION DATA:
45 APPLICATION NUMBER:  US 07/429,791
46 FILING DATE:  31-OCT-1989
47
48 ATTORNEY/AGENT INFORMATION:
49 NAME:  INGOLIA, DIANE E.
50 REGISTRATION NUMBER:  40,027
51 REFERENCE/DOCKET NUMBER:  OPND-01763
52 TELECOMMUNICATION INFORMATION:
53 TELEPHONE:  (415) 705-8410
54 TELEFAX:  (415) 397-8338
55 INFORMATION FOR SEQ ID NO. 27:
56 SEQUENCE CHARACTERISTICS:
57 LENGTH:  3891 base pairs
58 TYPE:  nucleic acid
59 STRANDEDNESS:  double
60 TOPOLOGY:  linear
61 MOLECULE TYPE:  DNA (genomic)
62 FEATURE:
63 NAME/KEY:  CDS
64 LOCATION:  1...3888
65
66 US-08-480-604A-27

```

Db 2572 CTACGTATTGGAAAGCTGAAGGTATCTGTACTAAGTGTAT 2527

Search completed: September 16, 2002, 03:26:45  
Job time: 9328 sec

Query Match	9.5%;	Score 30;	DB 1;	Length 3891;
Best Local	Similarly	48.8%;	Pred. No. 11;	
Matches	81;	Conservative	0;	Mismatches 85; Indels 0; Gaps 0;

QY 104 tttagggagacatacaagaagtcagttggttggtgccttgttgagtrcccaactatattcaagt 163  
 Db 2692 TTTTGTAGCATCCTAGATTAAGTCTATTAAAGATTAACTTCATATCTTAATTAACATA 2633  
 QY 164 agatatacaccacaatttgaataatgatgtgcacacaataaacgaggttaatgaaac 223  
 Db 2632 TAGAAGATTAATTAATATTTTAAATATATTCACGTAATATAGATTAATATCTTGATTAT 2573  
 QY 224 atcagtagaagaatacacacattccatcccttaagaagatcatc 269

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 01:42:43 : Search time 3013.46 Seconds  
(without alignments)  
1410.850 Million cell updates/sec

Title: US-09-700-770-4  
Perfect score: 315  
Sequence: 1 taacactgactcagattt.....tggtattactactatgc 315

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

27472414

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_hic:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_hiv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	313.4	99.5	563	10	BF335657 RCI-CT046
2	132.4	42.0	513	10	MA9446 yw22a12.r1
3	70.8	22.5	323	9	AW753759 RCI-CT028
4	41.8	13.3	844	12	CNS000FR AL069961 Drosophila
5	40	12.7	952	12	CNS01016 AL098388 Drosophila
6	40	12.7	1101	12	CNS00DT7 AL075293 Drosophila
7	39.6	12.6	652	10	BF295719 Drosophila
8	39.2	12.4	1204	12	CNS016E2 BF295719 027BFE02
9	39	12.4	634	9	AL639578 AL106628 Drosophila
10	39	12.4	718	10	BF572670 AL639578
11	38.8	12.3	463	10	BI360054 602079265
12	38.8	12.3	650	10	BI360054 384881 MA
13	38.4	12.2	551	9	BE291792 AW746159
14	37.8	11.9	1225	12	CNS016ID AW746159 WSL_39.DO
15	37.4	11.9	698	12	AL106171 Drosophila
16	37.2	11.8	417	12	BH421278 BH421278 BOHCQ62TR
17	37.2	11.8	588	12	AZ784456 BM538650 BOHPV56TR

18	36.8	11.7	563	9	AL587852
19	36.6	11.6	564	12	AO457327
20	36.6	11.6	601	9	AL643997
21	36.4	11.6	394	12	AO66673 HS_2195_B
22	36.4	11.6	1086	12	CNS00YXK AL096962 Drosophila
23	36.2	11.5	404	12	AO178095 HS_2217_B
24	36.2	11.5	421	10	BM162279 EST564802
25	36.2	11.5	1041	12	CNS015ID
26	36	11.4	806	12	CNS008OW
27	36	11.4	815	12	BH561753
28	35.8	11.4	517	12	B69927
29	35.8	11.4	520	12	BH062502
30	35.8	11.4	655	12	AG046736
31	35.8	11.4	690	12	AZ292215
32	35.8	11.4	710	12	BH502667
33	35.8	11.4	791	9	AW242927
34	35.8	11.4	795	10	BG783502
35	35.8	11.4	960	12	B18793
36	35.8	11.4	987	12	CNS014PO
37	35.8	11.4	1096	12	CNS05LPI
38	35.6	11.3	552	12	AO562320
39	35.6	11.3	875	12	AZ692611
40	35.4	11.2	544	12	BH042090
41	35.4	11.2	657	9	BB213292
42	35.2	11.2	326	9	AJ285573
43	35.2	11.2	430	12	AO879467
44	35.2	11.2	604	12	AZ591623
45	35.2	11.2	979	12	CNS012JO

## ALIGNMENTS

RESULT 1	BF335657	563 bp	mrna	linear	EST 22-NOV-2000
LOCUS	RCI-CT0469-160800-011-e04	CT0469	Homo sapiens	cdna	mrna sequence.
DEFINITION	BF335657				
ACCESSION	BF335657.1	GI:11306405			
VERSION					
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 563) Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Balda,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.				
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed				
JOURNAL	sequence tags				
MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)				
COMMENT	20202663 Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?l=RCI-CT0469-160800-011-e04&f=2000-08-16&t=1) Seq primer: puc 18 forward High quality sequence start: 21 High quality sequence stop: 559. Location/Qualifiers 1..563 /organism="Homo sapiens"				









QY	4	acacgactcagattttaaagaataactcttggagaatagagaacaatgaatcacttct	63
Db	619	AAATGAAAAAAACTTATATAAAAAAGTTTCATATANTTCATATNATGGTAATCTTAT	560
QY	64	ccaccacttaagatctatctcttaagagatctcaagctccctcttagggagacatacaagtc	123
Db	559	CCAGCAATCATTAATNGTGTATATGATGCCATGTTTCTATGANCCTAACATTTCACGA	500
QY	124	agtgctgtccttgtttagtccaccatbatctaaagtatgtagtactacaatttc	183
Db	499	ACTTTTGATATTAATTTTGGAGAGTGGCTTGAATGCAACCTCTTAATTAATTTTTT	440
QY	184	aaatagatgtcacacataaactggag	212
Db	439	CGGCTAATATTTTATTTATCTATCTAGTCG	411
RESULT	8		
CNS016E2			
LOCUS			
DEFINITION	CNS016E2	1204 bp	DNA linear GSS 26-JUL-1999
	Drosophila melanogaster genome survey sequence T7 end of BAC		
	BACN15A12 of DrosBAC library from Drosophila melanogaster (fruit		
	fly), genomic survey sequence.		
ACCESSION	AL106628		
VERSION	ALI06628.1	GI:5622852	
KEYWORDS	GSS.		
SOURCE	fruit fly.		
ORGANISM	Drosophila melanogaster		
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;		
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
REFERENCE	1 (bases 1 to 1204)		
AUTHORS	Genoscope.		
TITLE	Direct Submission		
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :		
	BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr		
COMMENT	- Web : www.genoscope.cns.fr)		
	Determination of this BAC-end sequence was carried out as part of a		
	collaboration with the European Drosophila Genome Project (EDGP) -		
	http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC		
	library (Dros BAC) was made by Alain Billaud at CBP (Centre		
	d'Etude du Polymorphisme Humain) with funding provided by a MRC		
	project grant. The DNA was prepared from embryos by Alain Bucheton		
	and Genevieve Payan. It has been constructed in the vector		
	pbeloBAC11.		
FEATURES			
Source	Location/Qualifiers		
	1..1204		
	/organism="Drosophila melanogaster"		
	/plasmid="pbeloBAC11"		
	/db_xref="taxon:7227"		
	/clone_lib="DrosBAC"		
	/clone="BACN15A12"		
	/note="end : 77"		
BASE COUNT	298 a 172 c 106 g 316 t 312 others		
ORIGIN			
	Query Match	12.4%	Score 39.2; DB 12; Length 1204;
	Best Local Similarity	21.1%	Pred. No. 14;
	Matches	62; Conservative 116; Mismatches 115; Indels 1; Gaps 1;	
QY	17	ttaaagaataactcttggagaatagagaacaatgaatcacttctccacacacttaagt	76
Db	748	TTTTTTTGAATTTTNCCTTTTAAATTTTCTATTAATAAATAAATTTCTMAADAATTTA	807
QY	77	atattcttagagatctacagctccctttagggagacatacaaaactgctgttcct	136
Db	808	AAATCTTTTCCSCCSCCSCCSCCMMCTTTTCTCTTMMNNNNNNNNNTTTTNNNNMM	867
QY	137	ttgttgatctccacattatctcaagtatgtagtactacaatttgaataagatgttc	196
Db	868	TTNNNTTNNNNNNNNNNNNNTTTNNNTTTNNNTTTTNNNNNTTTNNNNNTTTTNT	926

Oy	197	aacacaataactgaggtttatggaaacatcsgtgaagaagaaatacaaatccaccst	256
Db	927	TTHMMNNMTTTTTTTTHHMTTMMNMNTMMNMMNMMNMMNMMNMMNMMNMMN	986
Oy	257	tacagaatcatctactgcacaactagaataattgcatagtcatctact	310
Db	987	MMMMTMMNMNMNTTTGTWTGTAATTHMMNMNMNMNTMMNMNTTNHTNHGTMMNM	1040
RESULT	9		
LOCUS	AL639578/c		
DEFINITION	AL639578 XGC-neurula Silurana tropicalis cDNA clone TNeu012m17 5'		
ACCESSION	AL639578		
VERSION	AL639578		
KEYWORDS	EST.		
SOURCE	AL639578..1 GI:16791709		
ORGANISM	western clawed frog. Silurana tropicalis Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Xenopodinae; Silurana. 1 (bases 1 to 634) Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J. Sanger Xenopus tropicalis EST project 2001 (10_2001)		
REFERENCE	Unpublished (2001)		
AUTHORS	Contact: Huckle E		
TITLE	Sanger Centre		
JOURNAL	Hinxton, Cambridgeshire, CB10 1SA, UK		
COMMENT	Email: trop@sanger.ac.uk Sanger Xenopus tropicalis EST project 2001 TROPICALIS_SEQUENCE.ID: TNeu012m17.sp6 Sequencing primer: SP6 This sequence is from a Xenopus Gene Collection (XGC) library constructed by Aaron M. Zorn. Location/Qualifiers 1..634 /organism="Silurana tropicalis" /db_xref="taxon:8364" /clone_1fb="TNeu012m17" /clone_1lb="XGC-neurula" /dev_stage="neurula" /lab_host="Escherichia coli DH10B" /note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA was oligo dT primed from 5ug of poly A+ RNA from neurula. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end." BASE COUNT 200 a 112 g 112 g 209 t 1 others ORIGIN		
Query Match	12.4%; Score 39; DB 9; Length 634;		
Best Local Similarity	50.8%; Pred. No. 15;		
Matches	93; Conservative 0; Mismatches 90; Indels 0; Gaps 0;		
Oy	39	aataaacacaagaatcaagtttcgccaccacttaagiatatcctctaagatciacagc	98
Db	502	ATTATTACAGACAAGAAGCAATCATTTAACCTTAATAATTAACCCAATAGGGCTGTCTAC	443
Oy	99	ctcccttagaggagacatacaaaagtcagttgttgctcttgttgagtcgccacttatatt	158
Db	442	CCCCAATAAAGGGGTATATATCTAGTAGTGCGATCAATACAAAGGTCGTATTATATT	383
Oy	159	caagtagtlatgactacaatttgyaaaaatgatgtcacacaataaactggagtttatt	218
Db	382	ACAGAGAAAAGGAATCATTTTAAAAATCGAATTATTGATTAATAAGAGCTATAG	323
Oy	219	gaa 221	
Db	322	GGA 320	

RESULT	10
BF572670	
LOCUS	
DEFINITION	BF572670 718 bp mRNA linear EST_12-DEC-2000 6020792265F1 NIH_MGC_62 Homo sapiens CDNA clone IMAGE:4233489 5 , mRNA sequence.
ACCESSION	BF572670
VERSION	BF572670.1 GI:11646382
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 718) NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> . National Institutes of Health. Mammalian Gene Collection (MGC) Unpublished (1999)
TITLE	Contact: Robert Strassberg, Ph.D.
JOURNAL	Email: cga@bsf.remall.nih.gov
COMMENT	Tissue Procurement: ATCC/DCTD/DTF CDNA Library Preparation: CLONETECH Laboratories, Inc. CDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LITCM1080 row: b column: 18 High quality sequence stop: 326.
FEATURES	1 location/Qualifiers

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4253489"
/clone_id="NH_MGC_62"
/tissue_type="melanotic melanoma, high MDR"
/lab_host="DH10B (T1 phage-resistant)"
/notes="Organ: skin; Vector: pDNR-LIB (Clontech); Site.1: SfiI (ggcgccctggcgc); Site.2: SfiI (ggccatcatggc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCGCATTTAGCC-3', and 3' adaptor sequence: 5'-ATTCAGAGCGCCGAGCGCCGACATG-dT(30)BN-3' (where B = A, C, G or G and N = A, C, G, or T). Average insert size 1.9-75 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."
```

[illegible]

RESULT	11		
BI360054			
LOCUS	BI360054	463 bp	mRNA linear
DEFINITION	384481 MARC 2PIG Sus scrofa cDNA 5',		EST 01-AUG-2001
ACCESSION	BI360054		
VERSION	BI360054.1	GI:15056082	

KEYWORDS EST.  
SOURCE pig.  
ORGANISM Sus scrofa  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eulhemia; Cetartiodactyla; Suidae; Sus.  
AUTHORS 1 (bases 1 to 463)  
Fehrenkrug,S.C., Fekking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,  
Stone,R.T., Haeldon,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.  
and Keeler,J.W.  
TITLE Design and use of two pooled tissue normalized cDNA libraries for  
EST discovery in swine  
JOURNAL Unpublished (2000)  
COMMENT Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smiththe@mail.marc.usda.gov  
Single pass sequencing. Bases called and alt trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 18  
and -mismatch 12 options.  
PCR primers  
FORWARD: AGGAACAGCATATGACCAT  
BACKWARD: GTTTCACATCAGCAGC  
Plate: 133 row: C column: 24  
Seq primer: ATTAGGAGACATATAG.  
FEATURES  
Source  
1..463  
Location/Qualifiers

```

BASE COUNT      176 a      42 c      85 g      160 t
ORIGIN
/dn_xref="taxon:9823"
/clone_1b="MARC_2P1G"
/issue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."

```

Query Match	Similarity	12.3%	Score 38.8	DB 10	Length 463
Best Local	Similarity 48.2%		Pred. No. 16		
Matches 109	Conservative 0		Mismatches 117	Indels 0	Gaps 0
QY	20	taagaataacttttgaagaatagacaacaagaaatcagttcttcacacacttaagtata	79		
DB	112	TAAGAAATATTTTATTTTAAATAAAGTAACGACCTTTATCTTAAACGCTAAATTACT	171		
QY	80	tctcttagagactcagaagcccttccttagggacatacaagtcagtttggtcccttg	139		
DB	172	GATTATATTTTGGCTGAGCTGACCTAAAGTTCAAAATTTGTATGCGAAGTTTATTTT	231		
QY	140	tttagtccacactatattcaagtagtatagtactacaatttgaatatagattgcaca	199		
DB	232	TTCAGCTTCAAAAGCCTTCTTAAATAAAGTCTTTTACTCATGTGAAAAA	291		
QY	200	cataaactgaggttatatgsgaaacatcagtagaaggaataacacata	245		
DB	292	AA	337		

[illegible]

REFERENCE 1 (bases 1 to 650)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabs-r@mail.nih.gov](mailto:cgabs-r@mail.nih.gov)  
 Tissue Procurement: Jeffrey Green M.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LAM8555 row: k column: 23  
 High quality sequence stop: 564.

## FEATURES

source

Location/Qualifiers

1..650  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3499150"  
 /clone\_lib="NCI\_CGAP\_Mam6"  
 /sex="female, virgin"  
 /tissue\_type="infiltrating ductal carcinoma"  
 /dev\_stage="5 months"  
 /lab\_host="DH10B"  
 /note="Organ: mammary. Vector: pCMV-Sport6; Site\_1: Salt;  
 Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
 Library constructed by Life Technologies. Investigator  
 providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 193 a 98 c 185 g 174 t

ORIGIN

Query Match 12.3%; Score 38.8; DB 10; Length 650;  
 Best Local Similarity 56.2%; Pred. No. 17;  
 Matches 73; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 10 acccaatttaagaataacttttagaataagaacaatgaatcaagtcttcacca 69  
 Db 174 ATTCCACCTTTGGATGATCCGACGATCTTCAGATGAATGCTTTCTGCATA 115  
 QY 70 ctttaataatactcttagagatctacagcctccctttaagggaacataaagtcagt 129  
 Db 114 CTGAACATACCTCAATATACCTTACCCACACACGCTTGATCAACCATTCAGTGT 55  
 QY 130 gtgcctcttg 139  
 Db 54 CTTGGCCTTG 45

RESULT 13  
 AM746159 551 bp mRNA linear EST 19-JUL-2000  
 LOCUS AM746159/c

DEFINITION WS1\_39.D05.g1\_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,

ACCESSION AM746159

VERSION AM746159

KEYWORDS EST.

SOURCE

ORGANISM

Sorghum

Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 551)

AUTHORS Cordonnier-Pratt, M.-M., Gingle, A., Marsala, C., Sudman, M. and Pratt

L.H.

TITLE An EST database from Sorghum: water-stressed plants

JOURNAL Unpublished (2000)

COMMENT Contact: Cordonnier-Pratt MM

Department of Botany

The University of Georgia

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860  
 Fax: 706 542 1805  
 Email: [mmpratt@uga.edu](mailto:mmpratt@uga.edu)  
 Sequences have been trimmed to exclude PolyA, vector and regions  
 below Phred quality 16. The threshold for highest quality sequence  
 is 20.  
 Seq primer: T7  
 High quality sequence start: 42  
 High quality sequence stop: 551  
 POLYA=yes.

## FEATURES

source

Location/Qualifiers

1..551  
 /organism="Sorghum bicolor"  
 /db\_xref="taxon:4558"  
 /clone\_lib="Water-stressed 1 (WS1)"  
 /note="Organ: Mix of 5-week old plants on days 7 & 8 after  
 water was withheld; Vector: lambda Zap; Site\_1: XhoI;  
 Site\_2: EcoRI; The library was made from polyA RNA in the  
 cloning vector lambda Zap II. Clones to be sequenced were  
 prepared by mass excision."

BASE COUNT 144 a 107 c 125 g 175 t

ORIGIN

Query Match 12.2%; Score 38.4; DB 9; Length 551;  
 Best Local Similarity 55.1%; Pred. No. 21;  
 Matches 75; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 176 aaatttgaataagatgttcacacataaactgagttatggaacatcagtaagg 235  
 Db 528 AACTAGTAAGTAAGTAATTGGCTTACATCAACAGACCTTGACCAATGTACAATCAGG 469  
 QY 236 aaatcacacttcaccccttacagagatcattcttgaactcaggttaattgtca 255  
 Db 468 AGATCTCACCTCCACAGCTGTAGAGCTCCCGATATTTCAGTGACGAAAAATTGCA 409  
 QY 296 tgtgtattatcactt 311  
 Db 408 TATTAATATTCACAT 393

RESULT 14  
 CDS0161D 1225 bp DNA linear GSS 26-JUL-1999  
 LOCUS CDS0161D/c

DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC

BACN15C18 of Drosophila library from Drosophila melanogaster (fruit

fly), genomic survey sequence.

ACCESSION AL106171

VERSION AL106171

KEYWORDS GSS.

SOURCE

ORGANISM

Drosophila

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidae; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1225)

Genoscope.

Direct Submission

Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : [seqref@genoscope.cns.fr](mailto:seqref@genoscope.cns.fr))

- Web : <http://www.genoscope.cns.fr>

determination of this BAC-end sequence was carried out as part of a

collaboration with the European Drosophila Genome Project (EDGP) -

<http://www.edgp.ebi.ac.uk> . This Drosophila melanogaster BAC

library (Dros BAC) was made by Alain Billaud at CEPH (Centre

d'Etude du Polymorphisme Humain) with funding provided by a MRC

project grant. The DNA was prepared from embryos by Alain Bucheton

and Genevieve Payan. It has been constructed in the vector

pBelobAC11.

## FEATURES

source

Location/Qualifiers

1..1225  
 /organism="Drosophila melanogaster"  
 /plasmid="pBelobAC11"

BASE COUNT	ORIGIN	266 a	128 c	38 g	368 t	425 others
/db_xref="taxon:71227"						
/clone_lib="prosbac"						
/clone="BACN15C18"						
/note="end : Sp6"						

Query Match	12.0%	Score 37.8;	DB 12,	Length 1225;
Best Local Similarity	35.5%	Pred. No. 30;		
Matches 102; Conservative	48;	Mismatches 135;	Indels 2;	Gaps 1;

Qy 27 taacttttggagaatagacaacaaatgaaatcgaatttcgccacccttaagtataatcccta 86  
| : : || :|| : :|| : | : | : | : ||  
Db 1210 TTWTKATTTTWWAWAWAWAWAWAAAAWMTTWTATTAAATAATATAWATWTATTA 115

QY 87 gagatctacagccctcctttagggagcatacaagtcaattglttgcttgttgatgc 146  
       : : : | :| :||| :| :| : : :  
Db 1150 TTTTAAAWMTTTTTATATWATTWTAAATAAAAAAATATTAAWWMAWMAAW--AAWA 1093

QY 147 ccaccttatatccaagtagratgactacacaattttgnaaatgatgtgcacacaataaa 206  
 : : :::::  
 Db 1092 WAWAAWAAWTWTTWKAAATATAAAAATTAWATATATAAAAAATWAWWWWTWAAAAAATTAA 103

QY 207 CTGGAGCTCCTGGAACATCAGTAGAAGGAAATACAACATCCATCCCTTACAGAGATC 266  
| |||: | : | | | : : | | | |  
Db 1032 TTTTTTTTWTTTTANAAAAAAAATAATTAATTATWATATAAAWWTTAAAAAATA 973

972 AWTATTTAAWAWTTTTTTRAAWWTWTTTTTTTTWTTTAWAAW 926

RESULT 15  
BH421278

LOCUS	BH421278	698 bp	DNA	linear	GSS 12-DEC-2001
DEFINITION	BOHC62TR	BOHC	Brassicla oleracea	genomic clone BOHC62,	DNA sequence.

VERSION	BH421278.1	GI:17607006
KEYWORDS	GSS.	
SOURCE	Brassica oleracea.	

ORGANISM  
Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidales; eurosids II; Brassicales; Brassicaceae; Brassica

REFERENCE	AUTHORS	TITLE
1 (bases 1 to 698)	Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.	Whole genome shotgun sequencing of <i>Brassica oleracea</i>

COMMENT: Other\_GSSS: BOHCQ62  
Contact: Chris Town  
TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208

DNA is from a doubled haploid provided by Tom Osborn  
 Seq primer: TR  
 Class: sheared ends.

FEATURES	Location/Qualifiers
source	1. .698

```
/organism="Brassica oleracea"  
/strain="T01000DH3"  
/db_xref="taxon:3712"
```

```

/clone="BOHC062"
/clone_lib="BOHC"
/note="Vector: pHS01; Site:1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHS01 using BstXI linker"

```

BASE COUNT  
ORIGIN

Query Match	11.9%	Score 37.4	DB 12	Length 698
Best Local Similarity	48.8%	Pred. No. 36		
Matches 101	Conservative	0	Mismatches 106	Indels 0
				Gaps 0

**Qy**    24 aaataacttttgagaatagacaacaaatgaatcagtlcttcaccacttaagtatarctc 83  
       | | | | | | | | | | | | | | | | | | | | |  
**Db**    167 AGAACAATTTTAAAAAACTAAAGAAGAGCTAAAATTTCAGTTAAAAACCTTTTGTTCAAGAC 226

QY 84 tttagagatctacagcctccctttagggacatacaaaagtcagtltgttgctttgtta 143  
+ + + + + + + + + + + + + + + +  
Db 227 TCTGAAGCCTTCGGCTAACCGCTATCAGAGAACCATAAGACCCCTGACTTTTACCTTGTTCA 286

27 144 gtccaccattatattcaagtagtgatgactacaattttgaaatagatctgcacacaat 203  
 28 | | | | | | | | | | | | | | | | | | | |  
 Db 287 AGACGATCCTTGAATCTTGAGACTAATTTCTCCCATATCGAACCCATCTTCGATACAAA 346

|    |     |                              |     |
|----|-----|------------------------------|-----|
| QY | 204 | aaactcgagcttcatgaaacatcagta  | 230 |
|    |     |                              |     |
| Db | 347 | GCAATTAAACATAAGTTTACCCTCAGTA | 373 |

```
Search completed: September 16, 2002, 01:42:48
Job time: 3301 sec
```

|    |     |                             |     |
|----|-----|-----------------------------|-----|
| Qy | 204 | aaactgagtttatgtaaacatcagta  | 230 |
|    |     |                             |     |
| Db | 347 | GCAATTAAACATAAGTTTACCTCAGTA | 373 |

Search completed: September 16, 2002, 01:42:48  
 Job time: 3201.803

— 26 —

---

**THIS PAGE BLANK (UPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 06:24:21 : Search time 90.91 Seconds  
(without alignments)  
2418.238 Million cell updates/sec

Title: US-09-700-770-5

Perfect score: 895

Sequence: 1 ctaatcgttactgaacagc.....ataataattttaatgac 895

Scoring table:

OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents\_NA: \*  
1: /cgn2\_6/ptodata/2/lna/5A.COMB.seq: \*  
2: /cgn2\_6/ptodata/2/lna/5B.COMB.seq: \*  
3: /cgn2\_6/ptodata/2/lna/6A.COMB.seq: \*  
4: /cgn2\_6/ptodata/2/lna/6B.COMB.seq: \*  
5: /cgn2\_6/ptodata/2/lna/PTCUTS.COMB.seq: \*  
6: /cgn2\_6/ptodata/2/lna/Backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description       |
|------------|-------|-------------|--------|-------|-------------------|
| 1          | 18    | 2.0         | 774    | 4     | US-08-998-416-655 |
| 2          | 17    | 1.9         | 1503   | 3     | US-08-889-841B-1  |
| 3          | 17    | 1.9         | 1503   | 3     | US-08-889-841B-4  |
| 4          | 17    | 1.9         | 1545   | 1     | US-08-400-275-17  |
| 5          | 17    | 1.9         | 1990   | 1     | US-08-256-130A-1  |
| 6          | 17    | 1.9         | 2059   | 1     | US-08-256-130A-2  |
| 7          | 17    | 1.9         | 7498   | 2     | US-08-816-693A-1  |
| 8          | 17    | 1.9         | 7498   | 2     | US-08-885-291-1   |
| 9          | 17    | 1.9         | 7498   | 4     | US-09-496-672-1   |
| 10         | 16    | 1.8         | 275    | 1     | US-08-215-084A-2  |
| 11         | 16    | 1.8         | 275    | 1     | US-08-463-212-2   |
| 12         | 16    | 1.8         | 275    | 1     | US-08-463-212-2   |
| 13         | 16    | 1.8         | 318    | 3     | US-09-188-930-15  |
| 14         | 16    | 1.8         | 429    | 1     | US-08-215-084A-6  |
| 15         | 16    | 1.8         | 429    | 1     | US-08-463-212-6   |
| 16         | 16    | 1.8         | 429    | 1     | US-08-463-211-6   |
| 17         | 16    | 1.8         | 900    | 4     | US-09-085-371-5   |
| 18         | 16    | 1.8         | 1333   | 3     | US-08-889-425-3   |
| 19         | 16    | 1.8         | 1403   | 1     | US-07-872-678A-1  |
| 20         | 16    | 1.8         | 1601   | 1     | US-07-968-971A-10 |
| 21         | 16    | 1.8         | 1601   | 1     | US-08-424-406-6   |
| 22         | 16    | 1.8         | 1621   | 1     | US-07-937-609-22  |
| 23         | 16    | 1.8         | 1621   | 4     | US-08-029-170-22  |
| 24         | 16    | 1.8         | 1744   | 1     | US-08-424-406-7   |
| 25         | 16    | 1.8         | 1744   | 1     | US-08-464-523B-7  |
| 26         | 16    | 1.8         | 1810   | 1     | US-08-299-849B-20 |
| 27         | 16    | 1.8         | 1810   | 2     | US-08-142-368A-20 |

|    |    |     |      |   |                   |                    |
|----|----|-----|------|---|-------------------|--------------------|
| 28 | 16 | 1.8 | 1810 | 3 | US-08-967-727-20  | Sequence 20, Appl  |
| 29 | 16 | 1.8 | 1810 | 4 | US-08-037-230D-20 | Sequence 20, Appl  |
| 30 | 16 | 1.8 | 1850 | 1 | US-08-225-473-3   | Sequence 3, Appl1  |
| 31 | 16 | 1.8 | 2192 | 3 | US-08-942-001-1   | Sequence 1, Appl1  |
| 32 | 16 | 1.8 | 2192 | 4 | US-09-337-386-1   | Sequence 1, Appl1  |
| 33 | 16 | 1.8 | 2238 | 1 | US-07-841-651-1   | Sequence 1, Appl1  |
| 34 | 16 | 1.8 | 2542 | 1 | US-08-441-370-1   | Sequence 1, Appl1  |
| 35 | 16 | 1.8 | 2991 | 3 | US-08-795-430-48  | Sequence 48, Appl1 |
| 36 | 16 | 1.8 | 3060 | 1 | US-08-098-141-1   | Sequence 1, Appl1  |
| 37 | 16 | 1.8 | 3131 | 3 | US-09-035-648-23  | Sequence 23, Appl1 |
| 38 | 16 | 1.8 | 3131 | 4 | US-09-001-951-23  | Sequence 23, Appl1 |
| 39 | 16 | 1.8 | 3839 | 4 | US-09-056-105-14  | Sequence 14, Appl1 |
| 40 | 16 | 1.8 | 4060 | 1 | US-08-308-949A-1  | Sequence 1, Appl1  |
| 41 | 16 | 1.8 | 4584 | 2 | US-08-901-200A-15 | Sequence 15, Appl1 |
| 42 | 16 | 1.8 | 4584 | 2 | US-09-219-391-15  | Sequence 15, Appl1 |
| 43 | 16 | 1.8 | 5183 | 2 | US-08-870-518-7   | Sequence 7, Appl1  |
| 44 | 16 | 1.8 | 6418 | 1 | US-08-480-528A-11 | Sequence 11, Appl1 |
| 45 | 16 | 1.8 | 6418 | 1 | US-08-479-666-11  | Sequence 11, Appl1 |

#### ALIGNMENTS

RESULT 1  
US-08-998-416-655/c  
Sequence 655, Application US/08998416  
Patent No. 6239264  
GENERAL INFORMATION:  
APPLICANT: Philippsen, Peter  
APPLICANT: Pohlmann, Rainer  
APPLICANT: Steiner, Sabine  
APPLICANT: Mohr, Christine  
APPLICANT: Wendland, Jurgen  
APPLICANT: Knechtel, Philipp  
APPLICANT: Reibschung, Corinne  
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPTII  
NUMBER OF INVENTIONS: AND USES THEREOF  
NUMBER OF SEQUENCES: 1152  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6239264artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: No. 6239264th Carolina  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/998,416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 655:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 774 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: PAG1438UP

US-08-998-416-655

## Query Match

2.0%: Score 18; DB 4; Length 774;  
Best Local Similarity 100.0%; Pred. No. 9;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;QY 215 cccctccctgtgtgtc 222  
|||||  
Db 667 CCCCTCCTCTGTGTCTGC 650

## RESULT 2

US-08-889-841B-1  
; Sequence 1, Application US/08889841B  
; GENERAL INFORMATION:  
; APPLICANT: Bertman, Phillip W.  
; TITLE OF INVENTION: HIV ENVELOPE POLYPEPTIDES AND VACCINE  
; FILE REFERENCE: 14918-703C1P  
; CURRENT APPLICATION NUMBER: US/08/889,841B  
; CURRENT FILING DATE: 1997-07-08  
; PRIOR APPLICATION NUMBER: US 60/676,737  
; PRIOR FILING DATE: 1996-07-08  
; NUMBER OF SEQ. ID NOS: 57  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ. ID NO 1  
; LENGTH: 1503  
; TYPE: DNA  
; ORGANISM: HIV  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1503)  
US-08-889-841B-1

## Query Match

1.9%: Score 17; DB 3; Length 1503;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;QY 508 caattctcaacaatg 524  
|||||  
Db 765 caattctcaacaatg 781

## RESULT 3

US-08-889-841B-4  
; Sequence 4, Application US/08889841B  
; GENERAL INFORMATION:  
; APPLICANT: Bertman, Phillip W.  
; TITLE OF INVENTION: HIV ENVELOPE POLYPEPTIDES AND VACCINE  
; FILE REFERENCE: 14918-703C1P  
; CURRENT APPLICATION NUMBER: US/08/889,841B  
; CURRENT FILING DATE: 1997-07-08  
; PRIOR APPLICATION NUMBER: US 60/676,737  
; PRIOR FILING DATE: 1996-07-08  
; NUMBER OF SEQ. ID NOS: 57  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ. ID NO 4  
; LENGTH: 1503  
; TYPE: DNA  
; ORGANISM: HIV  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1503)  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1503)  
; OTHER INFORMATION: n = A,T,C or G  
US-08-889-841B-4Query Match  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;QY 508 caattctcaacaatg 524  
|||||  
Db 765 caattctcaacaatg 781

## RESULT 4

US-08-400-275-17  
; Sequence 17, Application US/08400275  
; Patent No. 5668295  
; GENERAL INFORMATION:  
; APPLICANT: Mahab, Samir Z.  
; APPLICANT: Malik, Vedpal S.  
; TITLE OF INVENTION: PUTRESCINE N-METHYLTRANSFERASE,  
; TITLE OF INVENTION: RECOMBINANT DNA MOLECULES ENCODING PUTRESCINE  
; TITLE OF INVENTION: N-METHYLTRANSFERASE, AND TRANSGENIC TOBACCO PLANTS WITH  
; TITLE OF INVENTION: ALTERED NICOTINE CONTENT  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Neave  
; STREET: 1251 Ave. of the Americas  
; City: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/400,275  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/076,681  
; FILING DATE:  
; APPLICATION NUMBER: US 07/613,160  
; FILING DATE: 14-NOV-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Creason, Gary L.  
; REGISTRATION NUMBER: 34,310  
; REFERENCE/DOCKET NUMBER: PM-1696  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-596-9000  
; TELEFAX: 212-596-9090  
; INFORMATION FOR SEQ. ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1545 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
US-08-400-275-17Query Match  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;QY 844 aatgcaatcagaagaag 860  
|||||  
Db 1227 AATGCAATCAGGAAG 1243

## RESULT 5

US-08-256-130A-1  
; Sequence 1, Application US/08256130A  
; Patent No. 5659121  
; GENERAL INFORMATION:  
; APPLICANT: BIRD, COLIN R.

APPLICANT: GRIERSON, DONALD  
APPLICANT: HALL, LISA N.  
TITLE OF INVENTION: DNA, DNA CONSTRUCTS, CELLS AND PLANTS  
TITLE OF INVENTION: DERIVED THEREFROM  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.  
STREET: 1100 New York Avenue, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005-3918  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/256,130A  
FILING DATE: 30-SEP-1994  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9127098.3  
FILING DATE: 20-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT.GB92.02355  
FILING DATE: 18-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: KOKULIS, PAUL N.  
REGISTRATION NUMBER: 16,773  
REFERENCE/DOCKET NUMBER: 205427/SEE 36697/UST  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1990 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-256-130A-1

Query Match 1.9%; Score 17; DB 1; Length 1990;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 710 ttgtgtgtgtctct 726  
|||||  
Db 102 TTTGTTGTGCTCT 118

RESULT 6  
US-08-256-130A-2  
; Sequence 2, Application US/08256130A  
; Patent No. 5659121  
; GENERAL INFORMATION:  
; APPLICANT: BIRD, COLIN R.  
; APPLICANT: GRIERSON, DONALD  
; APPLICANT: HALL, LISA N.  
; TITLE OF INVENTION: DNA, DNA CONSTRUCTS, CELLS AND PLANTS  
; TITLE OF INVENTION: DERIVED THEREFROM  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.  
; STREET: 1100 New York Avenue, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3918  
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/256,130A  
FILING DATE: 30-SEP-1994  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9127098.3  
FILING DATE: 20-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT.GB92.02355  
FILING DATE: 18-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: KOKULIS, PAUL N.  
REGISTRATION NUMBER: 16,773  
REFERENCE/DOCKET NUMBER: 205427/SEE 36697/UST  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2059 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-256-130A-2

Query Match 1.9%; Score 17; DB 1; Length 2059;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 710 ttgtgtgtgtctct 726  
|||||  
Db 99 TTTGTTGTGCTCT 115

RESULT 7  
US-08-816-693A-1/C  
; Sequence 1, Application US/08816693A  
; Patent No. 5874241  
; GENERAL INFORMATION:  
; APPLICANT: Takahashi, Joseph S  
; APPLICANT: Turek, Fred W  
; APPLICANT: Pinto, Lawrence H  
; TITLE OF INVENTION: Clock Gene and Gene Product  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dressler, Rocky, Milnamow & Katz  
; STREET: Two Prudential Plaza, Suite 4700  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/816,693A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 5874241thrup, Thomas E  
; REGISTRATION NUMBER: 33,268  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-616-5400  
; TELEFAX: 312-616-5460

; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7498 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 389..2954  
; US-08-816-693A-1

Query Match 1.9%; Score 17; DB 2; Length 7498;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 ctcaaatgggaacgctg 66  
|||||  
Db 7255 CTCGAATGGGAACGCTG 7239

RESULT 8  
US-08-885-291-1/c  
; Sequence 1, Application US/08885291A  
; Patent No. 6057125  
; GENERAL INFORMATION:  
; APPLICANT: Takahashi, Joseph S.  
; APPLICANT: Turek, Fred W.  
; TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT  
; FILE REFERENCE: 0290-5  
; CURRENT APPLICATION NUMBER: US/08/885,291A  
; EARLIER FILING DATE: 1997-06-30  
; EARLIER APPLICATION NUMBER: 08/816,693  
; NUMBER OF SEQ ID NOS: 55  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 7498  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; US-08-885-291-1

Query Match 1.9%; Score 17; DB 3; Length 7498;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 ctcaaatgggaacgctg 66  
|||||  
Db 7255 CTCGAATGGGAACGCTG 7239

RESULT 9  
US-09-496-672-1/c  
; Sequence 1, Application US/09496672  
; Patent No. 6291429  
; GENERAL INFORMATION:  
; APPLICANT: Takahashi, Joseph S.  
; APPLICANT: Turek, Fred W.  
; APPLICANT: Pinto, Lawrence H.  
; TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT  
; FILE REFERENCE: 0290-5  
; CURRENT APPLICATION NUMBER: US/09/496,672  
; CURRENT FILING DATE: 2000-02-03  
; PRIOR APPLICATION NUMBER: 08/885,291  
; PRIOR FILING DATE: 1997-06-30  
; PRIOR APPLICATION NUMBER: 08/816,693  
; PRIOR FILING DATE: 1997-03-13  
; NUMBER OF SEQ ID NOS: 55  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1

; LENGTH: 7498  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; US-09-496-672-1

Query Match 1.9%; Score 17; DB 4; Length 7498;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 ctcaaatgggaacgctg 66  
|||||  
Db 7255 CTCGAATGGGAACGCTG 7239

RESULT 10  
US-08-215-084A-2/c  
; Sequence 2, Application US/08215084A  
; Patent No. 5461032  
; GENERAL INFORMATION:  
; APPLICANT: Karen J. Krapcho, Bradford Carr Van  
; APPLICANT: Magenen, J.R. Hunter Jackson and Robert  
; APPLICANT: M. Kral, Jr.  
; TITLE OF INVENTION: INSECTICIDALLY EFFECTIVE  
; TITLE OF INVENTION: PEPTIDES  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz  
; ADDRESSEE: and No. 5461032rls  
; STREET: One Liberty Place - 46th Floor  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: U.S.A.  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/215,084A  
; FILING DATE: March 18, 1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 859,925  
; FILING DATE: March 24, 1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: John W. Caldwell, Esq.  
; REGISTRATION NUMBER: 28,937  
; REFERENCE/DOCKET NUMBER: FMC-0069  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-568-3100  
; TELEFAX: 215-568-3439  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 275 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; US-08-215-084A-2

Query Match 1.8%; Score 16; DB 1; Length 275;  
Best Local Similarity 100.0%; Pred. No. 92;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 268 aatccagggtgcgca 283  
|||||  
Db 115 AATCCAGGGGTGCGCA 100

RESULT 11  
US-08-463-212-2/c

```
; Sequence 2, Application US/08463212
; Patent No. 5658563
; GENERAL INFORMATION:
; APPLICANT: Karen J. Krapcho, John Randolph Hunter
; APPLICANT: Jackson, Bradford C. VanWagenen, Robert
; APPLICANT: M. Kral, Jr.
; TITLE OF INVENTION: INSECTICIDALLY EFFECTIVE
; TITLE OF INVENTION: PEPTIDES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz
; ADDRESSEE: and No. 5658563rls
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 diskette, 720 Kb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,212
; FILING DATE: Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/215,084
; FILING DATE: 18-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: John W. Caldwell, Esq.
; REGISTRATION NUMBER: 28,937
; REFERENCE/DOCKET NUMBER: FWC-0076
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 275 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-08-463-212-2
;
Query Match 1.8%; Score 16; DB 1; Length 275;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 268 aatccagggtcgcca 283
DB 115 AATCCAGGGGTCCCA 100

RESULT 12
US-08-463-211-2/c
; Sequence 2, Application US/08463211
; Patent No. 5658781
; GENERAL INFORMATION:
; APPLICANT: Karen J. Krapcho, John Randolph Hunter
; APPLICANT: Jackson, Bradford C. VanWagenen, Robert
; APPLICANT: M. Kral, Jr.
; TITLE OF INVENTION: INSECTICIDALLY EFFECTIVE
; TITLE OF INVENTION: PEPTIDES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz
; ADDRESSEE: and No. 5658781rls
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
```

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 diskette, 720 Kb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,211
; FILING DATE: Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/215,084
; FILING DATE: 18-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: John W. Caldwell, Esq.
; REGISTRATION NUMBER: 28,937
; REFERENCE/DOCKET NUMBER: FWC-0075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 275 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-08-463-211-2
```

```
Query Match 1.8%; Score 16; DB 1; Length 275;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 268 aatccagggtcgcca 283
DB 115 AATCCAGGGGTCCCA 100
```

```
RESULT 13
US-09-188-930-15
; Sequence 15, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 318
; TYPE: DNA
; ORGANISM: Rat
;
US-09-188-930-15
```

```
Query Match 1.8%; Score 16; DB 3; Length 318;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 351 caaagatgttctcgt 366
DB 24 caaagatgttctcgt 39
```

```
RESULT 14
US-08-215-084A-6/c
; Sequence 6, Application US/08215084A
```

```
; Patent No. 5461032
; GENERAL INFORMATION:
; APPLICANT: Karen J. Krapcho, Bradford Carr Van
; APPLICANT: Magen, J.R. Hunter Jackson and Robert
; APPLICANT: M. Kral, Jr.
; TITLE OF INVENTION: INSECTICIDALLY EFFECTIVE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/215,084A
; FILING DATE: March 18, 1994
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 859,925
; FILING DATE: March 24, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: John W. Caldwell, Esq.
; REGISTRATION NUMBER: 28,937
; REFERENCE/DOCKET NUMBER: FMC-0069
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-08-215-084A-6
```

```
Query Match 1.8%; Score 16; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 268 aatccagggtcgcca 283
|||||
Db 288 AATCCAGGCGTCGCCA 273
```

```
RESULT 15
US-08-463-212-6/c
; Sequence 6, Application US/08463212
; Patent No. 5658563
; GENERAL INFORMATION:
; APPLICANT: Karen J. Krapcho, John Randolph Hunter
; APPLICANT: Jackson, Bradford C. VanWageningen, Robert
; APPLICANT: M. Kral, Jr.
; TITLE OF INVENTION: INSECTICIDALLY EFFECTIVE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz
; STREET: and No. 5658563r1s
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: 3.5 diskette, 720 Kb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,212
; FILING DATE: Herewith
; CLASSIFICATION: 514
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/215,084
; FILING DATE: 18-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: John W. Caldwell, Esq.
; REGISTRATION NUMBER: 28,937
; REFERENCE/DOCKET NUMBER: FMC-0076
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-08-463-212-6
```

```
Query Match 1.8%; Score 16; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 268 aatccagggtcgcca 283
|||||
Db 288 AATCCAGGCGTCGCCA 273
```

```
Search completed: September 16, 2002, 06:24:30
Job time: 10643 sec
```



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 05:18:16 : Search time 3017.73 Seconds  
(without alignments)  
4002.934 Million cell updates/sec

Title: US-09-700-770-5

Perfect score: 895

Sequence: 1 ctactcgttactgtaacgc.....aataaatatttttaaatgctc 895

Scoring table:

OLIGO\_NUC  
Gapop 60.0, Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description        |
|------------|-------|-------------|--------|----|--------------------|
| 1          | 521   | 58.2        | 693    | 10 | BG536626 602566231 |
| 2          | 497   | 55.5        | 498    | 10 | BF196645 7180f03.x |
| 3          | 455   | 50.8        | 491    | 9  | AT340056 qn18h09.x |
| 4          | 426   | 47.6        | 426    | 9  | AA479746 zu35g03.s |
| 5          | 398   | 44.5        | 504    | 9  | AI809681 w7f6g06.x |
| 6          | 397   | 44.4        | 397    | 9  | AA481852 zva4f04.r |
| 7          | 355   | 39.7        | 477    | 9  | AT142875 ox64e06.s |
| 8          | 355   | 39.7        | 563    | 9  | BE148882 CM2-HT024 |
| 9          | 339   | 37.9        | 431    | 9  | AT079465 o23f05.s  |
| 10         | 334   | 37.3        | 820    | 10 | BI770944 603059769 |
| 11         | 316   | 35.3        | 419    | 9  | AA430487 zw23f06.s |
| 12         | 315   | 35.2        | 429    | 9  | AM298444 UI-H-BMO- |
| 13         | 297   | 33.2        | 429    | 9  | AA317141 EST19048  |
| 14         | 269   | 30.1        | 491    | 9  | AT830551 w51b02.x  |
| 15         | 267   | 29.8        | 270    | 9  | AT860310 w10i010.x |
| 16         | 267   | 29.8        | 461    | 9  | AT276107 q174h09.x |
| 17         | 266   | 29.7        | 356    | 9  | AM770737 h189g02.x |

|      |     |      |      |    |          |
|------|-----|------|------|----|----------|
| C 18 | 263 | 29.4 | 263  | 10 | BE645174 |
| C 19 | 263 | 29.4 | 263  | 10 | BF055365 |
| C 20 | 263 | 29.4 | 408  | 9  | AM183010 |
| C 21 | 262 | 29.3 | 667  | 9  | AW471176 |
| C 22 | 261 | 29.2 | 482  | 10 | BF061539 |
| C 23 | 257 | 28.7 | 326  | 10 | BF882906 |
| C 24 | 256 | 28.6 | 404  | 10 | BF745581 |
| C 25 | 254 | 28.4 | 255  | 9  | AT207071 |
| C 26 | 250 | 27.9 | 411  | 9  | AT368232 |
| C 27 | 238 | 26.6 | 342  | 9  | AA477092 |
| C 28 | 238 | 26.6 | 392  | 9  | AA433968 |
| C 29 | 234 | 26.1 | 251  | 10 | BF743524 |
| C 30 | 208 | 23.2 | 254  | 9  | AM389442 |
| C 31 | 207 | 23.1 | 409  | 9  | BE073132 |
| C 32 | 205 | 22.9 | 410  | 10 | BF745586 |
| C 33 | 191 | 21.3 | 271  | 9  | BE181362 |
| C 34 | 188 | 21.0 | 232  | 9  | AT474243 |
| C 35 | 181 | 20.2 | 422  | 10 | BF743581 |
| C 36 | 177 | 19.8 | 302  | 9  | AM770575 |
| C 37 | 159 | 17.8 | 223  | 9  | AM797203 |
| C 38 | 155 | 17.3 | 356  | 9  | AM173795 |
| C 39 | 135 | 15.1 | 144  | 9  | AA074399 |
| C 40 | 133 | 14.9 | 267  | 10 | BF743520 |
| C 41 | 130 | 14.5 | 279  | 9  | AT758852 |
| C 42 | 122 | 13.6 | 401  | 9  | AT276273 |
| C 43 | 117 | 13.1 | 192  | 10 | CO0521   |
| C 44 | 115 | 12.8 | 1279 | 10 | BE535485 |
| C 45 | 114 | 12.7 | 227  | 9  | AT680037 |

#### ALIGNMENTS

RESULT 1  
LOCUS BG536626  
DEFINITION 602566231F1 NIH\_MGC\_77 Homo sapiens CDNA clone IMAGE:4691118 5',  
mRNA sequence.  
ACCESSION BG536626  
VERSION BG536626.1 GI:13528159  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Tissue Procurement: CLONTECH Laboratories, Inc.  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNU)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNU at:  
http://image.llnl.gov  
Plate: L10CM1509 row: m column: 07  
High quality sequence stop: 611.  
Location/Qualifiers  
1..613  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4691118"  
/clone\_lib="NIH\_MGC\_77"  
/lab\_host="DH10B (TI phage-resistant)"  
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site:1:  
SfiI (ggcgcctcgcc); Site:2: SfiI (ggcattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCGAGAGCCGAGCGCCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.9

kb (range 0.5-4.0 kb). 12/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MGC Library."

BASE COUNT 165 a 176 c 132 g 140 t

ORIGIN

Query Match 58.2%; Score 521; DB 10; Length 613;  
Best Local Similarity 99.8%; Pred. No. 1.3e-260;  
Matches 571; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 140 gcaaggaaaccttcattatattcttaagaacttaccgtgcagcagcttgagatga 199  
DB 14 GCAAGGAACTTCCTATATATCTTCAAGCAACTTACAGTGCACCGACAGTTCCGATGA 73  
QY 200 aaattcaactctctcctctcctctgtgtgtgcactaatgctgattgctcctta 259  
DB 74 AAGTTCTAATCTCTCCCTCCTCCTGTTGCTGCCACTAATGCTGATGTCATGCTCTCTA 133  
QY 260 gcaagcctgaatccagaggtctgcagagagcagagcagagcagagccttctagagat 319  
DB 134 GCAAGCTGAATCCAGGGGTGCGCAGAGGCCACAGGAGCCAGGCTTCTAGAGAT 193  
QY 320 ggtctccaggaagcgcccaagaatgtgagtgcagaagattggttctctgagagcccgagaa 379  
DB 194 GGTCTCCAGGAGGGCGGCAAGATGTGATGTCAGAAAGATTGCTTCTGAGAGCCCGAGAA 253  
QY 380 gaaattcatgacagtgctcgggtgcagaaagacagtgctcctctgctgattcattcaag 439  
DB 254 GAAATTCATGACAGTGTCTGGGCTGCCAAGAGACAGTCCCTGTGATCATTTTCAGG 313  
QY 440 gcaatgtgaagaaacaagacaccacaagcagacagaaagcacaagaacattccagag 499  
DB 314 GCAATGTGAAGAAACACAGACACCAAGGACACACAGAAAGCCAAACATTCACAGAG 373  
QY 500 cctggcagcaattcttcaacaatgtcagtaagaagcttctgctcctttagagagc 539  
DB 374 CCGTCCAGCAATTTCTCAACAAATGTGATGAAGACTTGGCTTCTGTTGAGAGAC 433  
QY 560 tctgagcgccactctctcaattcaacattctcagcacaagaagacagtgagcacaactac 619  
DB 434 TCTGAGGGCCACTCTTCCATTAACATTTCTCAGCCAAAGACAGTGTGACACACTTAC 493  
QY 620 cagacactcttcttccacactcactcctcactgctacccaccttaactcactcagt 679  
DB 494 CAGACACTCTTCTTCCCTCCACTCCTCCTCCACTGTACCCACCCCTTAATTCATTCAGT 553  
QY 680 gctctcaaaaagatgttttcaagatcatt 711  
DB 554 GCTCTCAAAAAGCATGTTTTCAGATCATTT 585

RESULT 2  
BF196945/c 498 bp mRNA linear EST 03-NOV-2000  
LOCUS 7180f03.x1 Soares\_NSF\_F8\_9W\_OT\_PA\_P\_S1 Homo sapiens cDNA clone  
DEFINITION IMAGE:3529/32 3', mRNA sequence.  
ACCESSION BF196945  
VERSION BF196945.1 GI:11085527  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 498)  
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
Tumor Gene Index  
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
COMMENT Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.jnl.gov) for further information.  
High quality sequence stop: 487.  
Location/Qualifiers

FEATURES  
Source

1.498  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3527932"  
/clone\_lib="Soares\_NSF\_F8\_9W\_OT\_PA\_P\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pUT73P-Pac (Pharmacia) with a modified polylinker; Site:1: Not 1; Site:2: Eco RI; Equal amounts of plasmid DNA from five normalized libraries were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 5 libraries. The pools consisted of the following libraries and clones: Soares NBHP pool 1: 309384-310919, 323208-325895 Soares NBHP pool 1: 145032-147335, 147720-148103, 148872-149255, 15002 - 150407, 151176-152327 Soares NBHPF8-9W pool 1: 758280-760583, 772104-774407 Soares NBHP pool 1: 304776-306311, 320136-322823, 326280-326663 Soares NBHP pool 1: 723720-726407, 739080-740999 Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 132 a 83 c 138 g 145 t

ORIGIN

Query Match 55.5%; Score 497; DB 10; Length 498;  
Best Local Similarity 100.0%; Pred. No. 4.3e-248;  
Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 399 tgggtctgccaagaagcagtgccctctgctatcttcaagggaatgtgaagaagaag 458  
DB 498 TGGGCTGCCAAGAAAGCAGTGCCTCTGTGATCTTTCAAGGCAATGTGAAGAAACAG 439  
QY 459 acaccaagcagcacagaagaagcacaagcattccagagcctgcccagcaatttccaa 518  
DB 438 ACACCAAGGACACACAGAAAGCCAAAGCATTCACAGGCTGCGACGACATTTCTCA 379  
QY 519 acaatgtcagtaagaagcttctctctgtgtgtagagctctgagccactcttc 578  
DB 378 ACAATGTCACTAAGAAAGCTTGTCTGCTTGTGTGAGAGCTCTGAGCGCCACTCTTCC 319  
QY 579 aattaaacattctcagcagaagaagacagtgagcaccctcagacactcttctcc 638  
DB 318 AATTAAACATTTCTCAGCAGAAAGACAGTACGACACACTTCTTCTTCTCC 259  
QY 639 acctcaactctccactgtaccacccctaaatcattccagtgctcctcaaaaagcatgtt 698  
DB 258 ACCTCACTCTCCACAGTACCCACCCCTTAATCATTCAGTCTCAAAAACCATGTTT 199  
QY 699 ttcaagatcatttgtttgttctctctctctctctctctctctctctctctctctctct 758  
DB 198 TTCAAGATCATTTTGTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 139  
QY 759 gtgcccctccctaccagcagtgtaacttaattaccctgaagaagcttcaggaactctagc 818  
DB 138 GTGCCCTCCCTTACCCAGGCTTAGGCTTAATTAAGCTGAAGATTCACGAAACCTGTAC 79  
QY 819 ttccctagctagtgctaatcaacttaattgaatgcaagaagtagcaagaagctcaat 878  
DB 78 TTCCCTAGCTAGTGTCAATTAAAGCATTCAGAAAGATGCAAGACAGAAAGTCAAT 19  
QY 879 aaatatttttaaatgtc 895  
DB 18 AATATATTTTAATATGTC 2

RESULT 3  
A1340056/c 491 bp mRNA linear EST 13-FEB-1999  
LOCUS A1340056

| DEFINITION  | ACCESSION | VERSION  | KEYWORDS | SOURCE | ORGANISM     | REFERENCE  | AUTHORS            | TITLE  | JOURNAL   | COMMENT   |
|---|-----------|----------|----------|--------|--------------|--|--------------------|--|---|---|
| q9180h09.x1 NCI-CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1898657 3' | U1340056  | U1340056 | EST.     | human. | Homo sapiens | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | 1 (bases 1 to 491) | NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> . | National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index   | Unpublished (1997)  |
|   |           |          |          |        |              |  |                    |  | Contact: Robert Strausberg, Ph.D.<br>Email: cgaps@email.nih.gov<br>Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmerl-Buck, M.D., Ph.D. | CDNA Library Preparation: M. Bento Soares, Ph.D.<br>CDNA Library Arrayed by: Greg Lennon, Ph.D.<br>DNA Sequencing by: Washington University Genome Sequencing Center<br>Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at:<br><a href="http://www.bio.linnl.gov/bdpp/image/image.html">www.bio.linnl.gov/bdpp/image/image.html</a><br>Insert length: 485 Std Error: 0.00<br>Seq primer: ~40UP from Gibco<br>High quality sequence stop: 444. |
|   |           |          |          |        |              |  |                    |  | Location/Qualifiers   | 1..491  |

|    |     |  |     |
|----|-----|--|-----|
| Yy | 519 | acaaatgtaagctaaagaagctttgctctgaccttctgtaagagctctgaagcgcacactcttc   | 578 |
| Db | 251 | ACACCAAAAGGACACACAGAAAGCCAAACAGCATTTCCAGACCTTGCACGACATTTTCTAA      | 192 |
| Qy | 459 | acacccaagaagcaccagaaaagcacaacaagaacattccagagccctgcagcagaatttccaa   | 518 |
| Db | 311 | TGGGCTGTGCCAAAGAAAGCAGTGCCTCCTGTGATCATTTCAAGGGCATGTGAAGAAACAG      | 252 |
| Qy | 399 | tgagctgcacaaagaagcagtgccctcgtgatacatlttcaaggagcaatgtgaagaaacaag    | 458 |
| Db | 371 | AGATGTGTAGTGCAAAGATTGTTCTCTAGAGAGCCCCGAGAAAGAAATTCATGACAGTTC       | 312 |
| Qy | 339 | agaatgtgaatgcaaaagattggttccttgaagagcccgagaagaataatcatgacagatgctc   | 398 |
| Db | 431 | CGCCAGAGGCCACAGGGGACCGAGGCCAGGCTTTCAGGAGATGCGTCCAGGAAGCGCGCCA      | 372 |
| Qy | 279 | cgccagaagccacaagagcagcgagccagagctcttaagaagatgctctcaagaaagcagcgca   | 338 |
| Db | 491 | CCTCTGTGTGCTGGCACATATGCTGATGTCCATGTGCTCTTAGCAGCCCTGAAATTCAGGGGT    | 432 |
| Qy | 219 | cctccgtgtgctgcacacatgctcgaatgcatgctcagatgctctagcagcctgaaatccagaggt | 278 |
| Db | 491 | CCTCTGTGTGCTGGCACATATGCTGATGTCCATGTGCTCTTAGCAGCCCTGAAATTCAGGGGT    | 432 |

Query Match 50.8%; Score 455; DB 9; Length 491;

Best Local Similarity 100.0%; Prid. No. 3,6e-226;

Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

BASE COUNT 101 a 107 c 141 g 142 t

ORIGIN

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:1898657"

/clone\_lib="NCI-CGAP\_Lu5"

/lissue\_type="carcinoid"

/lab\_host="DH10B"

/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from neuroendocrine lung carcinoid, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

|    |     |   |      |
|----|-----|---|------|
| Dd | 191 | ACATGTCAGCTAAGAAGACTTTGGCTTGCCTTTGTAGSAGCTCTTGAGGCCCACTCTTCC    | 1338 |
| Oy | 579 | aattaacatcttcacagccaaagaagacagtagacacactaccagacaactccttctccg    | 636  |
| Dd | 131 | AATTAAACATTTCTCAGCCCAAGAAGACAATGTGACACACTTACCAGACACTCTTCTTCTCCC | 72   |
| Oy | 639 | aactcaactctccacatgtlaccaccaccccctaatacat                        | 673  |
| Dd | 71  | ACCTCACTCTCCCATGCTGACCCACCCCCAATAATCAT                          | 37   |

|             |   |   |        |      |        |                 |
|-------------|---|---|--------|------|--------|-----------------|
| RESULT      | 4 | AA479746/c                                      | 426 bp | mrna | linear | EST 08-AUG-1997 |
| LOCUS       |   | AA479746  |        |      |        |                 |
| DEFINITION  |   | 235593.61 Soares ovary tumor NbhOT Homo sapiens |        |      |        |                 |
| DESCRIPTION |   | IMAGE:740020 3', mRNA sequence.                 |        |      |        |                 |
| ACCESSION   |   | IMAGE:740020                                    |        |      |        |                 |

|           |              |            |
|-----------|--------------|------------|
| ACCESSION | AA4179740    |            |
| VERSION   | AA4179746.1  | GI:2205632 |
| KEYWORDS  | EST.         |            |
| SOURCE    | human.       |            |
| ORGANISM  | Homo sapiens |            |

REFERENCE  
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 426)  
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,

TITLE  
JOURNAL  
COMMENT

Washo-Merck EST Project 1997  
Unpublished (1997)  
Contact: Wilson RK

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [est@watson.wustl.edu](mailto:est@watson.wustl.edu)  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.lln.gov](mailto:info@image.lln.gov)) for further information.  
Seq primer: -41m3 fwd. ET from Amersham  
High quality sequence stop: 419.

|                           | FEATURES   |                     |
|---------------------------|--|---------------------|
|                           | SOURCE   | Location/Qualifiers |
|                           |  | 1..426              |
|                           | /organism="Homo sapiens"   |                     |
|                           | /db_xref="GDB:593972"  |                     |
|                           | /db_xref="taxon:9606"  |                     |
|                           | /clone_image="IMAGE:740020"  |                     |
|                           | /clone_lib="Soares ovary tumor NbHOT"  |                     |
|                           | /sex="Female"  |                     |
|                           | /tissue_type="Ovarian tumor"   |                     |
|                           | /lab_host="DH10B (ampicillin resistant)"   |                     |
|                           | /note="Organ: ovary; Vector: pPVTSD (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'-TGATACCAATCTGAAGTAGGGCGGCCGCCTTTTTTTTTTTTTT 3']". double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pRT3 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo." |                     |
| BASE COUNT                | 121 a      64 C      122 g      119 t  |                     |
| ORIGIN                    |  |                     |
| Query Match               | 47.6%; Score 426; DB 9; Length 426;  |                     |
| Best Local Similarity     | 100.0%; Pred. No. 4.7e-211;  |                     |
| Matches 426; Conservative | 0; Mismatches 0; Indels 0; Gaps 0;   |                     |
| Oy                        | 469 cacacagaagccaaacaagcatcgcgagctcgctcgaacatttcataaatgtcag  | 528                 |
|                           |  |                     |
| Dd                        | 426 CACCACAGAAGCCAAACAAGCATTCAGAGCCTCCACAGCAATTCTCAACAATGTGAG  | 367                 |
| Oy                        | 529 cttaagaagcttgctcgtcctttgtagagactctgagcccaactctccaattaacat  | 588                 |
|                           |  |                     |

Db 366 CTAGAAGCTTTGCTGCTGCTTTGTAGAGCTCTGAGCGCCCACTCTTCCATTAAACAT 307  
 QY 589 tctaacgaagaagacagtagcaccacacacacacactctctctccacacactctc 648  
 Db 306 TCTCAGCGAAGAGAGAGTAGCAGCAGCTACACAGACTCTCTCTCCACCTCAGCTCT 247  
 QY 649 cccactgtacccaccccttaataatcctcagtgctcctcaaaaagcagcttttcaaga 708  
 Db 246 CCCACTGTACCCACCCCTTAATCATTCAGTGTCTCAAAAAGCATGTCTTTTCAAGATCA 187  
 QY 709 ttttgtttgtgtctctcctcagtgctctctctctcgtcagtcctcagtcgtccccc 768  
 Db 186 TTTTGTGTGTGCTCTCTCTAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 127  
 QY 769 cttaaccagcttagccttaataatcctcgaagaattccaggaactcagtcctcagctc 828  
 Db 126 CTTACCCAGGCTTAGGCTTAACTTACCTGAAGAAGATCCAGGAAGACTGTAGCTCTCT 67  
 QY 829 gtgtcatttaaccttaaatgcaatcagaagaagtagcaagaagctcaataatatttt 888  
 Db 66 GTGTCAATTAACTTAATGCAATCAGGAAGTAGCAAAAGTCAATTAATATTTTTT 7  
 QY 889 aaatgt 894  
 Db 6 AAATGT 1  
 RESULT 5 504 bp mRNA linear EST 07-JUL-1999  
 AI809681  
 LOCUS AI809681/c  
 DEFINITION wf76c06.x1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
 IMAGE:2361514 3', mRNA sequence.  
 ACCESSION AI809681  
 VERSION AI809681.1 GI:5396247  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 504)  
 NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 Seq primer: -40UP from Gihco  
 High quality sequence stop: 459.  
 FEATURES  
 source  
 1..504  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2361514"  
 /clone\_1lb="Soares\_NFL\_T\_GBC\_S1"  
 /lab\_host="DH10B"  
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with  
 a modified polylinker. Site.1: Not I; Site.2: Eco RI;  
 Equal amounts of plasmid DNA from three normalized  
 libraries (fetal lung NBH19W, testis NHT, and B-cell  
 NCI-CGAP-GCB1) were mixed, and ss circles were made in  
 vitro. Following HAP purification, this DNA was used as  
 tracer in a subtractive hybridization reaction. The driver  
 was PCR-amplified cDNAs from pools of 5,000 clones made  
 from the same 3 libraries. The pools consisted of  
 1.M.A.G.E. clones 297480-302087, 682632-687239,  
 726408-728711, and 729096-731399. Subtraction by Bento  
 Soares and M. Fatima Bonaldo. "  
 BASE COUNT 134 a 85 c 141 g 144 t  
 ORIGIN

Query Match 44.5%; Score 398; DB 9; Length 504;  
 Best Local Similarity 99.8%; Pred. No. 2e-196;  
 Matches 448; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 407 caaagaagcagtgccctgtgacatcattcaagggcaatgtgaagaagaagaccacaa 466  
 Db 490 CAAAGAAGCAGTGCCCTGTGATCATTTCAAGGCAATGTGAAGAAACAAAGACACCAA 431  
 QY 467 ggcaccacagaagaagcccaaacattccagagcctcgcagcaattctcaacaatgic 526  
 Db 430 GGCACCCAGAAAGCCAGAGACATTCAGAGCTGCGCAGCATTTTCAACAATGTC 371  
 QY 527 agctaaagaagcttgctcgtccttctgaagagctctgaagccacactctccaattaac 586  
 Db 370 AGCTAAGAAGCTTTGGCTGCTCTCTTGTAGAGAGCTGAGGCGCCACCTTCCAAATTAAC 311  
 QY 587 atctcagccaagaagaagcagtagcagacacactaccagaacactctctctccacact 646  
 Db 310 ATTCTCAGCCAGAAAGACAGTGAAGACACCTACAGACACTCTCTCTCTCCACCTCACT 251  
 QY 647 ctcccaactgtaccacccccttaataatcctcagtgctcctcaaaaagcagatgtttcaaat 706  
 Db 250 CTCCCACTGTACCCACCCCTTAATCATTCAGAGCTCTCAAAAAGCATGTTTCAAGAT 191  
 QY 707 catttgtttgtgtctctcctcagtgctcctctctcgtcagtcctcagtcgtgctc 766  
 Db 190 CATTTGTGTGTGCTCTCTCTAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 131  
 QY 767 ccccttaccagcgttagccttaataatcctcgaagaagatccagaagaactgtagctctc 826  
 Db 130 CCTTACCCAGGCTTAGGCTTAATTAATCCTGAAGATTCAGGAAGCTGTAGCTCTTCT 71  
 QY 827 tagtgtcatttaaccttaaatgcaatcag 855  
 Db 70 TAGTGTCAATTAACTTAATGCAATCAG 42  
 RESULT 6  
 AA481852 397 bp mRNA linear EST 08-AUG-1997  
 LOCUS AA481852  
 DEFINITION zva2f04.r1 Soares ovary tumor NBH07 Homo sapiens cDNA clone  
 IMAGE:756319 5', mRNA sequence.  
 ACCESSION AA481852  
 VERSION AA481852.1 GI:2209530  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 397)  
 Hillier, L., Allen, M., Bowles, L., Dubuque, T., Giesel, G., Jost, S.,  
 Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,  
 Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wyllie,  
 T., Waterston, R. and Wilson, R.  
 WashU-Merck EST Project 1997  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Wilison RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 Seq primer: -28mJ3 rev2 ET from Amersham.  
 FEATURES  
 source  
 1..397  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:756319"  
 /clone\_1lb="Soares ovary tumor NBH07"  
 /sex="Female"  
 /tissue\_type="Ovarian tumor"

```

/lab host="DH10B (ampicillin resistant)"
/ncb="organ: ovary; Vector: p7773 (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5
TGTTACCAATCTGAAGTGGAGCGCGCCGGTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p7773 vector
(Pharmacia). Library constructed by Bento Soares and
M.Fatima Bonaldo."

```

|                       |              |           |            |            |
|-----------------------|--------------|-----------|------------|------------|
| Query Match           | 44.4%        | Score 397 | DB 9       | length 397 |
| Best Local Similarity | 100.0%       | Pred. No. | 6.4e-196   |            |
| Matches 397           | Conservative | 0         | Mismatches | 0          |
|                       |              |           | Indels     | 0          |
|                       |              |           | Gaps       | 0          |

| QY  | 445 |
|---|-----|
| ggaagaagaacagacacccaagagcaccacagaagaagcccaacagcattccagaagccgc | 504 |
|   |     |
| Db  | 1   |
| GTGAAGAAACAGACACCAAAGGCACACAGAAAGCCAAACAGCATTCCAGAGCCTGC      | 60  |

|    |     |  |     |
|----|-----|--|-----|
| Oy | 505 | cagcaatttcacaacatgctcagctaagaagcttgcctcgctttgtatgaagctcttga  | 564 |
|    |     |  |     |
|    |     |  |     |
|    |     |  |     |
| Db | 61  | CAGCAATTTCTAACAATGTCTAGCTAAAGAAGTTCCTGCTTGCTTTGTAAGGACCTCTGA | 120 |

| Accession | Sequence  | Position |
|-----------|---|----------|
| Oy        | gagcccaactcttcgaatlaaaccttcacgacagaagaacagtagagcaccctccacagac | 624      |
| Db        | 121 GGGCCCACTCTTCGAATTAACATTCACGCCAAGAACAAGTAGCACCCTACACAGAC  | 180      |

|    |     |   |     |
|----|-----|---|-----|
| Oy | 625 | acttcttcctccacctacacbtcccaactgtaacccaacccctaaatacatctccagtgccct   | 684 |
|    |     |   |     |
|    |     |   |     |
|    |     |   |     |
| Db | 181 | actcttcttctccaccctcacctctgccactgtatccaccacccctaaatcatctccagtgctct | 240 |

Oy 685 caaaagcatgtttttcaagatcalcttgtttgttcctctctctagtcctctctctc 744  
 Db 241 caaaacatgtttttcaagatcatcttggctgctctctctctctctctc 300

**DY**    745   gtcagcttagcctgcygcctccccctaaccagagctatggcttaattacacctgaagaattc    804  
**Db**       301   gtcagctttagcctgtgccctcccttaaccacggcttgaggctaatgatccctgaagaattc    360

| QY | 805 | caggaactgttagcttcctagctagtgtcaattaac | 841 |
|----|-----|--------------------------------------|-----|
| Db | 361 | CAGGAACTGTAGCTTCCTAGCTAGTGTCAATTAA   | 397 |

RESULT 7  
A1142875/C

|            |  |
|------------|--|
| DEFINITION | ox64e06.s1 Soares_NhMMPu_S1 Homo sapiens CDNA clone IMAGE:1661122          |
| ACCESSION  | A1142875   |
| FEATURES   | <p>1..1000 bp linear</p> <p>1..1000 bp linear</p> <p>3' mRNA sequence.</p> |

|          |              |
|----------|--------------|
| ORGANISM | Homo sapiens |
| SOURCE   | human.       |
| KEYWORDS | EST.         |
| ENZYME   | GLYC2675.1   |
| PROTEIN  | GLYC2675.1   |

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NCBI-GAP <http://www.ncbi.nlm.nih.gov/ncbigap>

|          |                                      |
|----------|--------------------------------------|
| JOURNAL  | Tumor genome Anatomy Project (CGAP), |
| COMMENT  | National Cancer Institute, Cancer    |
| TITLE    | Tumor Gene Index                     |
|          | Unpublished (1997)                   |
| Contact: | Robert Strausberg                    |
|          | Ph D.                                |

Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert Length: 890 Std. Dev: 0

High quality sequence stop: 383.  
location/Qualifiers  
1..477

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:166112"
/clone_id="Scares_NbHMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10b"
/note="Organ: mixed (see below); Vector: pRTT30-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NbHMP, pregnant uterus
NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of 1.M.A.G.E. clones 260232-265223,

```

| BASE COUNT<br>ORIGIN | 129 a | 74 c | 135 g | 139 t |
|----------------------|-------|------|-------|-------|
|                      |       |      |       |       |

|                           |       |                    |           |             |
|---------------------------|-------|--------------------|-----------|-------------|
| Query Match               | 39.7% | Score 355;         | DB 9;     | Length 477; |
| Best Local Similarity     | 99.8% | Pred. NO.5.5e-174; |           |             |
| Matches 475; Conservative | 0;    | Mismatches 0;      | Indels 1; | Gaps 1;     |

|    |     |  |     |
|----|-----|--|-----|
| QY | 420 | ccccctgatacatcattcaaggccaatgtgaagaaacacagaccaccaagaacacccagaaa | 479 |
| Db | 477 | ccccttgatcatttttcnaagggcaatgtgaagaaaaaacagacaccacaacacaaaa     | 418 |

|    |     |   |     |
|----|-----|---|-----|
| Q7 | 480 | gccaaacaagcattccagagcctgcccagcaatttctcaaacatgtagctagaagctt  | 539 |
| Db | 417 | gccaaacaagcatttccagagcctgcccagcaatttctcaaacatgtagctagaagctt | 358 |

| QY | 540   | 599 |
|----|---|-----|
| QY | tgctcgaccttgtagagccttgagcgccacctcttccaatlaaaraatctcagccaag      | 599 |
| Db | 357 T-CTGTGCTTTGTAGAGAGCTCTGTAGCGCCCTCTTTCCAATTAATCATTCTCAGCAAG | 299 |

QY 600 aagacagtgagcacaccctaccagacactcttctctccacactcaactctccactgtacc 659

Db 298 AAGACAGTGAGCACACCCTACCAGACACTCTTCTTCTCCACACTCACTCTCCACTGTACC 239

660 cacccttaatcatctccagtgctctcaaaaagcatglttllccaagatcaatltgttctt 719  
QY  
238 CACCCTTAATCATCTCCAGTGCCTCAAAAAGCATGTTTTCGAAGATCATTTGTGTGT  
Dp  
179

OY 720 gctcctatagtcgtcttctctctatcgtaagtcttagccctgtagccctaccacgaac 779  
|||||  
Db 178 GGTCTCTCTAGATGTCCTTCTCTCTCGTAGTCTTAGCCTGTGCCCTCCCTTACCCAGG 119  
|||||

Oy 780 ttagcctaattacctaagaattccagaaactgtagcttcctagctagtgatcattaa 839  
 Db 118 tttagccttaattaccctgaagattccagaaactgtagcttcctagctagtgatcattaa 59

|     |   |     |
|-----|---|-----|
| 840 | cctlaaattgcaatcaggaagttagcaacagagtcaaataattttaattc  | 895 |
| QY  |   |     |
| 58  | ctttaaattgccaattcagcaaacagagctcattataatttttttaatttc |     |
| DH  |   |     |
|     | ctttaaattgccaattcagcaaacagagctcattataatttttttaatttc | 3   |

RESULT 8  
00140007

|            |                            |         |                    |                 |
|------------|----------------------------|---------|--------------------|-----------------|
| LOCUS      | 563 bp                     | mRNA    | linear             | EST 21 JUN-2000 |
| DEFINITION | CM2-HT02244-221199-038-f02 | HT02244 | Homo sapiens cDNA, | mRNA sequence.  |
| ACCESSION  | BE148882                   |         |                    |                 |
| VERSION    | BE148882.1                 |         |                    |                 |
| DATE       | 07-06-2000                 |         |                    |                 |

ORGANISM Homo sapiens  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS  
1 (bases 1 to 563)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R., Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F.



| RESULT | LOCUS    | DEFINITION  | ACCESSION   | VERSION             | KEYWORDS  | SOURCE          | ORGANISM |
|--------|----------|---|-------------|---------------------|-----------|-----------------|----------|
| 10     | B1770944 | B1770944<br>603059769p1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5209073 5', mRNA sequence.  | 820 bp      | mRNA                | linear    | EST 25-SEP-2001 |          |
|        |          | B1770944  |             |                     |           |                 |          |
|        |          | B1770944.1  | GI:15762522 |                     |           |                 |          |
|        |          | human.  |             |                     |           |                 |          |
|        |          | Homo sapiens  |             |                     |           |                 |          |
|        |          | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;   |             |                     |           |                 |          |
|        |          | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  |             |                     |           |                 |          |
|        |          | NIH-MGC http://mgc.nci.nih.gov/.<br>National Institutes of Health, Mammalian Gene Collection (MGC)  |             |                     |           |                 |          |
|        |          | Unpublished (1999)<br>Contact: Robert Strausberg, Ph.D.<br>Email: cgabs@email.nih.gov   |             |                     |           |                 |          |
|        |          | Tissue Procurement: Life Technologies, Inc.<br>CDNA Library Preparation: Life Technologies, Inc.<br>CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)   |             |                     |           |                 |          |
|        |          | DNA sequencing by: Incyte Genomics, Inc.<br>Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at:<br><a href="http://image.lnl.gov">http://image.lnl.gov</a>   |             |                     |           |                 |          |
|        |          | plate: LHAM1524 row: j column: 18<br>High quality sequence stop: 772.<br>Location/Qualifiers  |             |                     |           |                 |          |
|        |          | 1..820  |             |                     |           |                 |          |
|        |          | /organism="Homo sapiens"  |             |                     |           |                 |          |
|        |          | /db_xref="taxon:9606"   |             |                     |           |                 |          |
|        |          | /clone="IMAGE:5209073"  |             |                     |           |                 |          |
|        |          | /clone_id="NIH_MGC_122"   |             |                     |           |                 |          |
|        |          | /lab_host="DH10B"   |             |                     |           |                 |          |
|        |          | /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 24 week female lung, 16 week female spleen, and 20-22 week male spleens. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (In vitro). Research Genetics tracking code 020. Note: This is a NIH_MGC Library." |             |                     |           |                 |          |
|        |          | this is a NIH_MGC Library."   |             |                     |           |                 |          |
|        |          | BASE COUNT  | 225 a       | 237 c               | 162 g     | 196 t           |          |
|        |          | ORIGIN  |             |                     |           |                 |          |
|        |          | Query Match   | 37.3%       | Score 334;          | DB 10;    | Length 820;     |          |
|        |          | Best Local Similarity   | 100.0%;     | Pred. No. 5,7e-153; |           |                 |          |
|        |          | Matches 334; Conservative   | 0;          | Mismatches 0;       | Indels 0; | Gaps 0;         |          |
|        |          | 378 aagaattctgaagtggtctggtccgcaaaagacagtccctgtgatcatctcaa   | 437         |                     |           |                 |          |
|        |          |   |             |                     |           |                 |          |
|        |          | 270 AAGAAATTCATGACAGTGCTGGGCTGCCAAGAAGCAGTCCCTGTGATCATTTCA  | 329         |                     |           |                 |          |
|        |          |   |             |                     |           |                 |          |
|        |          | 438 ggccaatgtgaagaanaaacacacacaaaggcacacagaaagccaacaacattccag   | 497         |                     |           |                 |          |
|        |          |   |             |                     |           |                 |          |
|        |          | 330 GGGCAATGTGAAGAAACAAGACACCAAAGGACCACAGAAAGCCAACAACATTCCAG  | 389         |                     |           |                 |          |
|        |          |   |             |                     |           |                 |          |
|        |          | 498 agctctgcagcaaatcttccaacaatgtcagctaagaagcttgtctgtcgcttgtaga  | 557         |                     |           |                 |          |
|        |          |   |             |                     |           |                 |          |
|        |          | 390 AGCGCGCCAGCAAAATTTCCAAAACAAATGTGACGTAAAGAGCTTTCCTTTGTAGGA   | 449         |                     |           |                 |          |
|        |          |   |             |                     |           |                 |          |
|        |          | 558 gctctgagcggcccaactcttcacataaacattcttaagccaagaagaagtgagacact   | 617         |                     |           |                 |          |
|        |          |   |             |                     |           |                 |          |
|        |          | 450 GCTCTGAGCGGCCACTCTTCCATTAATAACATTCTTAGCCAAAGAACAAGTGGCACACT   | 509         |                     |           |                 |          |
|        |          |   |             |                     |           |                 |          |
|        |          | 618 accagagacctctctcccacacctcactctcccaacttacccacccccaaatcatcca  | 677         |                     |           |                 |          |
|        |          |   |             |                     |           |                 |          |
|        |          | 510 ACCAGACACTCTTCTTTCGCCACGTCACGTCCTGCCACTGTACCCACACCCGTAATCATTTCCA  | 569         |                     |           |                 |          |
|        |          |   |             |                     |           |                 |          |
|        |          | y , 678 gtgctctcaaaaaagcagtttlccaagatcattt 711  |             |                     |           |                 |          |

| Db                        | 570  | GTGCTCTCAAAAAGCATGTTTTTCAAGATCATTT | 603                         |
|---------------------------|--|------------------------------------|-----------------------------|
| RESULT 11                 | AAAA30487/c  | 419 bp                             | mRNA linear EST 16-OCT-1997 |
| LOCUS                     | AAAA30487  |                                    |                             |
| DEFINITION                | zw23106.s1 Soares ovary tumor NbHOT  | Human sapiens cDNA clone           |                             |
| ACCESSION                 | IMAGE:770147 3', mRNA sequence.  |                                    |                             |
| VERSION                   | AAAA30487  |                                    |                             |
| KEYWORDS                  | AAAA30487.1 GI:2111061   |                                    |                             |
| SOURCE                    | EST.   |                                    |                             |
| ORGANISM                  | human.   |                                    |                             |
| REFERENCE                 | Human sapiens  |                                    |                             |
| AUTHORS                   | Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 419)   |                                    |                             |
| TITLE                     | Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Kuwab, T., Lacy, M., Le, N., Lennon, G., Maita, M., Martin, D., Moore, B., Schell, M., Stepien, M., Tan, F., Theising, B., White, Y., Wyllie, T., Waterston, R., and Wilson, R.   |                                    |                             |
| JOURNAL                   | Washu-Merck EST Project 1997   |                                    |                             |
| COMMENT                   | Unpublished (1997)<br>Contact: Wilson RK<br>Washington University School of Medicine<br>4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108<br>Tel: 314 286 1800<br>Fax: 314 286 1810<br>Email: est@watson.wustl.edu<br>This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.<br>Seq primer: -41ml3 fwd. ET from Amersham<br>High quality sequence stop: 356.<br>Location/Qualifiers<br>1. 419<br>/organism="Homo sapiens"<br>/db_xref="GDB:5980022"<br>/db_xref="taxon:9606"<br>/clone="IMAGE:770147"<br>/clone_lib="Soares ovary tumor NbHOT"<br>/sex="Female"<br>/tissue_type="ovarian tumor"<br>/lab_host="DH10B (ampicillin resistant)"<br>/note="Organ: Ovary; Vector: p773D (Pharmacia) with a modified polylinker; Site: 1: Not I; Site: 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTACCAATCTGAAGTGGAGGAGCGCGGGTGTGTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified p773 vector (Pharmacia). Library constructed by Bento Soares and M. Fatima Bonaldo." |                                    |                             |
| BASE COUNT                | 121 a 63 c 117 g 118 t   |                                    |                             |
| ORIGIN                    |  |                                    |                             |
| Query Match               | 35.3%; Score 316; DB 9; Length 419;  |                                    |                             |
| Best Local Similarity     | 99.5%; Pred. No. 1.2e-153;   |                                    |                             |
| Matches 416; Conservative | 0; Mismatches 2; Indels 0; Gaps 0;   |                                    |                             |
| 478                       | aagccaacaagcattccagagcgtccagcaatttccaacaagtgcagctaaagaagc  | 537                                |                             |
| 479                       |  | 538                                |                             |
| 480                       |  | 539                                |                             |
| 481                       |  | 540                                |                             |
| 482                       |  | 541                                |                             |
| 483                       |  | 542                                |                             |
| 484                       |  | 543                                |                             |
| 485                       |  | 544                                |                             |
| 486                       |  | 545                                |                             |
| 487                       |  | 546                                |                             |
| 488                       |  | 547                                |                             |
| 489                       |  | 548                                |                             |
| 490                       |  | 549                                |                             |
| 491                       |  | 550                                |                             |
| 492                       |  | 551                                |                             |
| 493                       |  | 552                                |                             |
| 494                       |  | 553                                |                             |
| 495                       |  | 554                                |                             |
| 496                       |  | 555                                |                             |
| 497                       |  | 556                                |                             |
| 498                       |  | 557                                |                             |
| 499                       |  | 558                                |                             |
| 500                       |  | 559                                |                             |
| 501                       |  | 560                                |                             |
| 502                       |  | 561                                |                             |
| 503                       |  | 562                                |                             |
| 504                       |  | 563                                |                             |
| 505                       |  | 564                                |                             |
| 506                       |  | 565                                |                             |
| 507                       |  | 566                                |                             |
| 508                       |  | 567                                |                             |
| 509                       |  | 568                                |                             |
| 510                       |  | 569                                |                             |
| 511                       |  | 570                                |                             |
| 512                       |  | 571                                |                             |
| 513                       |  | 572                                |                             |
| 514                       |  | 573                                |                             |
| 515                       |  | 574                                |                             |
| 516                       |  | 575                                |                             |
| 517                       |  | 576                                |                             |
| 518                       |  | 577                                |                             |
| 519                       |  | 578                                |                             |
| 520                       |  | 579                                |                             |
| 521                       |  | 580                                |                             |
| 522                       |  | 581                                |                             |
| 523                       |  | 582                                |                             |
| 524                       |  | 583                                |                             |
| 525                       |  | 584                                |                             |
| 526                       |  | 585                                |                             |
| 527                       |  | 586                                |                             |
| 528                       |  | 587                                |                             |
| 529                       |  | 588                                |                             |
| 530                       |  | 589                                |                             |
| 531                       |  | 590                                |                             |
| 532                       |  | 591                                |                             |
| 533                       |  | 592                                |                             |
| 534                       |  | 593                                |                             |
| 535                       |  | 594                                |                             |
| 536                       |  | 595                                |                             |
| 537                       |  | 596                                |                             |
| 538                       |  | 597                                |                             |
| 539                       |  | 598                                |                             |
| 540                       |  | 599                                |                             |
| 541                       |  | 600                                |                             |
| 542                       |  | 601                                |                             |
| 543                       |  | 602                                |                             |
| 544                       |  | 603                                |                             |
| 545                       |  | 604                                |                             |
| 546                       |  | 605                                |                             |
| 547                       |  | 606                                |                             |
| 548                       |  |                                    |                             |

```

Db      239  CCCACCCCTTAATTCATTCAGTGTCTCTCAAAAGCATGTTTTCACAGATCATTTGTTG 180
Qy      718  ttgctctctagtgctctctctctctctctctctctctctctctctctctctctctctctctc
Db      179  TTGCTCTCTAGTGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
Qy      778  gcttaggcttaattactctgaagattccaggaacgcttagctctctctctctctctctctctc
Db      119  GCTTAGGCTTAATTCATTCAGTAAGATTCCTAGCTTCTCTAGCTAGTGTCTCTT 60
Qy      838  aaccttaagcaatcaggaagatgagcaaacagatgcaataatattttaaatgctc 895
Db      59  AACCTTAATTCATTCAGTAAGATTCAGTAAGATTCATTAATTTTAAATGTC 2

```

```

RESULT 12
AM298444/c 429 bp mRNA linear EST 16-JAN-2000
LOCUS      AM298444
DEFINITION UT-H-BW0-aj1-e-06-0-UT.s1 NCI_CGAP_Sub6 Homo sapiens cDNA clone
ACCESSION  AM298444
VERSION     AM298444.1 GI:6705080
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 429)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
             Email: cgaps-rt@mail.nih.gov
             The sequence contained an oligo-dT track that was present in the
             oligonucleotide that was used to prime the synthesis of first
             strand cDNA and therefore this may represent a bonafide poly A
             tail. cDNA library preparation: M.B. Soares Lab Clone distribution:
             NCI-CGAP clone distribution information can be found through the
             I.M.A.G.E. Consortium/ILNL at:
             www-bio.llnl.gov/dbirp/image/image.html
             Seq primer: M13 Forward
             POLYA=yes.

```

```

FEATURES
Source
Location/Qualifiers
1..429
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2732243"
/clone_lib="NCI CGAP_Sub6"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT773D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; NCI_CGAP_Sub6
is a subtracted library derived from BW, which consists of
a mixture of four normalized libraries: NCI_CGAP_Brn50,
NCI_CGAP_Lu13, NCI_CGAP_Ov18, GBC1. The NCI_CGAP_Sub6
library had 7 million recombinants. A single-stranded DNA
preparation of BW was used as a tracer in the IMAGE pool
hybridization with a driver comprising: the IMAGE pool
(NCI_CGAP_Kid3 pool 1 LAM 3334-3337, 3682-3683,
3798-3803 (IMAGE Clonoids 1322376-1323911,
1456008-1456775,1500552-1502855); NCI_CGAP_Kid5 pool 1
LAM 3338-3342,3722-3725, 3776-3778 (IMAGE Clonoids
1323912-1325831, 1471368-1472903, 1492104-1493255);
NCI_CGAP_Lu5 pool 1 LAM 3575-3582, 3851-3854 (IMAGE
Clonoids 1414920-1417991, 1520904-1522439); NCI_CGAP_GC4
pool 1 LAM 3164-3167, 3716-3720, 3733-3735 (IMAGE
Clonoids 1257096-1258631,1469064-1470983, 1475592-1476743
); NCI_CGAP_Pr22 pool 1 LAM 2457-2459, 2758-2759,
3062-3068 (IMAGE Clonoids 985608-986759, 1101192-1101959,
1217928-1220615); NCI_CGAP_Colo pool 1 LAM 2644-2653,
2871-2872 (IMAGE Clonoids 1057416-1061255,1144584-1145351
). (50% of the driver population), plus a pool of 3,840

```

```

arrayed clones from NCI_CGAP_Sub1 (IMAGE Clonoids
2708616-2710535) and NCI_CGAP_Sub2 (IMAGE Clonoids
2710536-2712455) (20% of the driver population), plus a
pool of 11,136 clones from NCI_CGAP_Sub3 (IMAGE Clonoids
2712456-2723591) (30% of the driver population).
Subtraction was performed as previously described [Bonaldo
, Lennon & Soares (1996): Normalization and Subtraction:
Two Approaches To Facilitate Gene Discovery. Genome
Research 6, 791-806.
TAG_LIB=NCI_CGAP-Lu13
TAG_TISSUE=lung
TAG_SEQ=GCCGG"
BASE COUNT 92 a 82 c 112 g 140 t 3 others
ORIGIN

```

```

Query Match 35.2%; Score 315; DB 9; Length 429;
Best Local Similarity 99.7%; Pred. No. 4,1e-153;
Matches 365; Conservative 0; Mismatches 1; Indels 0;

```

```

Qy      358  tggctctctgagagcccgaggaagaattcatgacagtgctctggtgccaagaagcag 417
Db      383  TGGTCTCTGAGAGCCCGAGAGAAATTCATGACGTGTCTGTGGCTGCCAAAGAGCAG 324
Qy      418  tgcctctgatacatctcgaaggcaatggaagaacaagacacacagcaccacaga 477
Db      323  TGCCCTGTGATCATTTTCAGAGGCAATGTGAGAAAACAAGACACCAAGGCCACAGA 264
Qy      478  aagccaaacagacatccagagcctgcccagcaattctcaaaaatgtaagctgaagaac 537
Db      263  AAGCCAAACAGACATTCACAGAGCCCTGCACAGATTTCTCAACAAATGTCAGTAAAGAG 204
Qy      538  ttgctctgctctgtagagagctctgagcgccacactctccaattaacattctcagca 597
Db      203  TTGCTCTGCTCTTGTGAGAGCTCTGAGCCGACACTCTTCATTAACATTCCTCAGCCA 144
Qy      598  agaagacagtgaagacacctaccagacactctctctccaccactacatctccacatgta 657
Db      143  AGAAGACAGTGAAGACACACNATCCAGACACTCTCTCTCCACCTCACTCTCCACTGTA 84
Qy      658  cccacccctaatacatctcagctgctctcctcaaaagcatgtttccaagatcattgttg 717
Db      83  CCCACCCCTTAATTCATTCAGTGTCTCTCAAAAGCATGTTTTCACAGATCATTTGTTG 24
Qy      718  ttgctc 723
Db      23  TTGCTC 18

```

```

RESULT 13
AA317141/c 297 bp mRNA linear EST 19-APR-1997
LOCUS      AA317141
DEFINITION EST19048 Lung Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION  AA317141
VERSION     AA317141.1 GI:1969489
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 297)
AUTHORS     Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
             ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
             ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Val,C., Clayton,R.A.,
             Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
             ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
             Gheini,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M.,
             Kelley,J.C., Liu,L.-T., Mamaros,S.M., Merrick,J.M.,
             Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
             Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R.,
             Small,K.V., Spriggs,T.A., Uterchag,T.R., Weidman,J.F., Li,Y.,
             Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
             Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.

```



| ORGANISM  |   |
|---|---|
| Homo sapiens  |   |
| Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |   |
| Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.         |   |
| 1 (bases 1 to 491)  |   |
| REFERENCE   |   |
| NCI-CCAP  | <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> . |

|            |   |
|------------|---|
| RESULT     | 15  |
| AIB60310/c |   |
| LOCUS      | 270 bp mRNA linear EST 26-AUG-1999  |
| DEFINITION | w101e10.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:242358 3',<br>mRNA sequence. |
| ACCESSION  | AIB60310  |
| VERSION    | AIB60310.1 GI:5513926   |
| KEYWORDS   | EST.  |

SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 270)  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgaps@email.nih.gov](mailto:cgaps@email.nih.gov)  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: [www-bio.lnl.gov/biopr/image/image.html](http://www-bio.lnl.gov/biopr/image/image.html)  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 262.  
 Location/Qualifiers  
 1..270  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2423658"  
 /clone\_lib="NCI CGAP Lnl9"  
 /tissue\_type="squamous cell carcinoma, poorly differentiated (4 pooled tumors, including primary and metastatic)"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pooled lung tumor tissue, and was then primed with a Not I - oligo(dt) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."  
 BASE COUNT 89 a 37 c 74 g 70 t  
 ORIGIN  
 Query Match 29.8%; Score 267; DB 9; Length 270;  
 Best Local Similarity 100.0%; Pred. No. 4, 2e-128;  
 Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 629 ttctctcccaactcactccacatgtaccacccctaaatcattccagtgctctcaaa 688  
 |||||||  
 DB 270 TTCTTCCCACTCCTCCTCCACCTGACCTGACCCCTTAATCATTCACAGTCTCAAA 211  
 QY 689 aagcagtgcttccaagatcattgttggctctctcctagtgctctctctcgta 748  
 |||||||  
 DB 210 AAGCATGTTTTCAGATCAATTTGTTGTCCTCTAGTCTCTCTCTCGTCA 151  
 QY 749 gtcttaagcctgtgcctccctaccacagcttaagcttaattacctgaagatccagg 808  
 |||||||  
 DB 150 GTCTTAGCCTGTGCGCTCCCTTACCAAGGCTTAGGCTTAATTACCTGAAGATTCAGG 91  
 QY 809 aaactgtagcttccctagctagtgatcattaaacttaaatgcaatcaggaagtagcaaac 868  
 |||||||  
 DB 90 AAAGTGTAGCTTCTGCTAGTCTAGTCAATTAATGCAATCAGAAAGTAGCAAC 31  
 QY 869 agaagtcataataatattttaaatgtc 895  
 |||||||  
 DB 30 AGAAGTCATATAATATTTTAAATGTC 4

Search completed: September 16, 2002, 05:18:19  
 Job time: 8392 sec



THIS PAGE BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 01:42:48 ; Search time 3013.46 Seconds  
(without alignments)  
4008.606 Million cell updates/sec

Title: US-09-700-770-5

Perfect score: 895  
Sequence: 1 ctaatcgtctacgacgc.....aataaatttttaaatgac 895

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estb:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description         |
|------------|-------|-------------|--------|----|---------------------|
| 1          | 672.2 | 75.1        | 820    | 10 | BT770944 603059769  |
| 2          | 582.6 | 65.1        | 613    | 10 | BG536626 60256231   |
| 3          | 510.6 | 57.1        | 667    | 9  | AW471176 xvi3a03.x  |
| 4          | 497   | 55.5        | 498    | 10 | BF196945            |
| 5          | 488.4 | 54.6        | 491    | 9  | AI340056 qn18h09.x  |
| 6          | 487.2 | 54.4        | 504    | 9  | AI809681 w76c06.x   |
| 7          | 471   | 52.6        | 491    | 9  | AI830551 w51b02.x   |
| 8          | 465.2 | 52.0        | 482    | 10 | BF061539 7j52910.x  |
| 9          | 464   | 51.8        | 477    | 9  | AI142875 ox64e06.s  |
| 10         | 431.8 | 48.2        | 461    | 9  | AI1276107 q174h09.x |
| 11         | 427.6 | 47.8        | 563    | 9  | BE148882 CM2-HT024  |
| 12         | 426   | 47.6        | 426    | 9  | AA479746 zuz3g03.s  |
| 13         | 414.8 | 46.3        | 419    | 9  | AA430487 zw23f06.s  |
| 14         | 409   | 45.7        | 429    | 9  | AM298444 UT-H-BW0   |
| 15         | 405   | 45.3        | 431    | 9  | AI079465 oz39f05.s  |
| 16         | 397   | 44.4        | 397    | 9  | AA481852 zv42f04.r  |
| 17         | 375.2 | 41.9        | 404    | 10 | BF745581 QVO-BT084  |

|    |       |      |      |    |                    |
|----|-------|------|------|----|--------------------|
| 18 | 365   | 40.8 | 392  | 9  | AA433968 zw23f07.r |
| 19 | 341.2 | 38.1 | 356  | 9  | AM770737 h189g02.x |
| 20 | 339.4 | 37.9 | 409  | 9  | BE073132 MR3-BT055 |
| 21 | 317.4 | 35.5 | 410  | 10 | BF745586 QVO-BT084 |
| 22 | 305.6 | 34.1 | 356  | 9  | AM173795 xp96910.x |
| 23 | 303   | 33.9 | 422  | 10 | BF743581 QVO-BT084 |
| 24 | 301.2 | 33.7 | 326  | 10 | BF882906 QVO-BT084 |
| 25 | 297   | 33.2 | 297  | 9  | AA317141 EST19048  |
| 26 | 287.6 | 32.1 | 302  | 9  | AM770575           |
| 27 | 270.4 | 30.2 | 408  | 9  | AM183010           |
| 28 | 269.4 | 30.1 | 387  | 10 | BM430624           |
| 29 | 269   | 29.1 | 275  | 9  | AA366288           |
| 30 | 267   | 29.8 | 270  | 9  | AI860310 w101e10.x |
| 31 | 264.8 | 29.6 | 411  | 9  | AI368232 qw16e02.x |
| 32 | 263   | 29.4 | 263  | 10 | BE645174 7e64d07.x |
| 33 | 263   | 29.4 | 263  | 10 | BF055365           |
| 34 | 258.4 | 29.0 | 479  | 10 | BE624584           |
| 35 | 258.6 | 28.9 | 605  | 9  | AA537274 vk46b02.r |
| 36 | 257.8 | 28.8 | 466  | 9  | AA881969 vx31g04.r |
| 37 | 257.8 | 28.8 | 484  | 9  | AA789593 wv61f07.r |
| 38 | 254   | 28.4 | 255  | 9  | AI207071 qf80f02.x |
| 39 | 253.4 | 28.3 | 433  | 9  | BE197231 ug73b06.y |
| 40 | 241.2 | 26.9 | 271  | 9  | BE181362 CM0-HT063 |
| 41 | 240.4 | 26.9 | 829  | 10 | BI078998 602873258 |
| 42 | 240.4 | 26.9 | 1279 | 10 | BG535485 602563170 |
| 43 | 238   | 26.6 | 242  | 9  | AA477092 zu35g03.r |
| 44 | 234   | 26.1 | 251  | 10 | BF743524 QVO-BT084 |
| 45 | 219   | 24.5 | 254  | 9  | AM389442 ILO-ST016 |

## ALIGNMENTS

RESULT 1  
LOCUS BT770944 820 bp mRNA linear EST 25-SEP-2001  
DEFINITION 603059769P1 NIH\_MGC\_122 Homo sapiens cDNA clone IMAGE:5209073 5',  
mRNA sequence.  
ACCESSION BT770944  
VERSION BT770944.1 GI:15762522  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE 1 (bases 1 to 820)  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1999)  
Contact: Robert Strusberg, Ph.D.  
Email: cga@pds-remail.nih.gov  
Tissue Procurement: Life Technologies, Inc.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at:  
http://image.llnl.gov  
Plate: LAM11524 row: j column: 18  
High quality sequence stop: 772.  
Location/Qualifiers  
1..820

## FEATURES

source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="5209073"  
/clone\_id="NIH\_MGC\_122"  
/lab\_host="DH10B"  
/note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;  
Site\_1: NotI; Site\_2: EcoRV (destroyed); RNA source  
anonymous pool of 24 week female lung, 16 week female  
spleen, and 20-22 week male spleens. Library is oligo-dT  
primed and directionally cloned (EcoRV site is destroyed  
upon cloning). Average insert size 1.4 kb, insert size

range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 026. Note: this is a NIH\_MGC library."

BASE COUNT 225 a 237 c 162 g 196 t

ORIGIN

Query Match 75.1%; Score 672.2; DB 10; Length 820;  
Best Local Similarity 98.4%; Pred. No. 2.8e-163;  
Matches 721; Conservative 0; Mismatches 8; Indels 4; Gaps 4;

QY 108 cctgagatcctcccaaggatcagagagcctcccaaggaaacctccatlatatctcaaa 167  
Db 1 CCTGAGTCATCCCA-GGATCAGAGAGCCCTCAGCAGGAACTTCATTTATCTTCA 59  
QY 168 gcaactaagagctgaccagagcagtgatgataaagttctatctctcctcctgtt 227  
Db 60 GCAACTTACAGCTCCACCGACAGTGGATGAAATTTATCTCTTCCCTCCCTGTT 119  
QY 228 gctgccaatagtctgaatgcatatgctctagaagccttaaccagaggtgcagag 287  
Db 120 GCTGCCATATGCTGATGTCATGTCATGCTCTACAGCCTTAATCCA-GGCTCCGACAGG 178  
QY 288 ccacagagagcagagcagagcctctcagagatgagctccagagcgcccaagaatgtga 347  
Db 179 CCACAGGACCGAGGCGGCTTCTAGAGATGGCTCCAGAGGCGCCACAGATGTGA 238  
QY 348 gtgcaaatgtgttctctg-agagcccgagaaagaatcagaagtgctcgtggtgc 406  
Db 239 GTGCAGAGATTTGTTCTCTGCTCGCCCCGCAAGAAATTCATACAGTCTGGGCTGC 298  
QY 407 caaagaagcagtgccctgtgatcatctcagaggcaatgtgaagaacaagaacacaa 466  
Db 299 CAAAGAAGCAGTGCCTGCTGATCATTTCAAGGCAATGTGAAGAAACAGACACCAA 358  
QY 467 ggcacacacaaagacaaagaacatccagaagcctgcagcaatcttctcaacaatgtc 526  
Db 359 GGCACACACAAAGCCAAACACATTCAGAGCTGCCGCAATTTCTCAACAAATGTC 418  
QY 527 agctaaagaagcttgctccttgtagagagctcgaagcgcacatcttccaatcaaa 586  
Db 419 AGCTAAGAAGCTTTCCTGCTGCTTTGTAGAGAGCTTGAGGCCACTTTCATTAATTAAC 478  
QY 587 attctcagcacaagaagcagtgagcaacactacagacaactcttcttccacactact 646  
Db 479 ATTCTAGCCAAAGACAGTGAAGACACCTACCAAGACACTTCTTCTGCCACTCAGT 538  
QY 647 ctccacatgacccaccccttaaatcattcgaagtcctcctcaaaaagatgttttcaaat 706  
Db 539 CTCCCACTGTACCCACCCCTAAATTCATTCAGTGTCTCAAAAAGCATGTTTTCAAGAT 598  
QY 707 catltgttgtgtctctctctagtgcttctctctctcgaagcttcaagcctgtgcctc 766  
Db 599 CA-TTGTGTTGGTGTCTCTGATGTCTTCTCTCTCTGATCTTAAAGCTGTGCTTC 657  
QY 767 ccttaaccagagcttagcttaatacctgaagaatccagaagaactgtagcttctcagc 826  
Db 658 CCTTACCCAGAGGCTTAGGCTTAATTAACCTGAAGATTCAGGAATGTAGCTTCTTAC 717  
QY 827 tagtgcatctaa 839  
Db 718 TAGTGTCAATTTTA 730

RESULT 2  
BG536626 613 bp mRNA linear EST 03-APR-2001  
LOCUS 602566231.1 NIH\_MGC\_77 Homo sapiens cDNA clone IMAGE:4691118 5',  
DEFINITION mRNA sequence.  
ACCESSION BG536626  
VERSION BG536626.1 GI:13528159  
KEYWORDS EST.

SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Euteria; Primates; Catarrhini; Hominiidae; Homo.  
TITLE 1 (bases 1 to 613)  
JOURNAL NIH-MGC http://imgc.nci.nih.gov/.  
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-r@mail.nih.gov  
Tissue Procurement: CLONTECH Laboratories, Inc.  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM1509 row: m column: 07  
High quality sequence stop: 611.  
Location/Qualifiers  
1. 613  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4691118"  
/clone\_lib="NIH\_MGC\_77"  
/lab\_host="DH10B (TI phage-resistant)"  
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site: 1: S11 (ggcgctcgcc); Site: 2: S11 (ggcattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTAATGACC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGGAGCGGCGGACGATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.9 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC library."

BASE COUNT 165 a 176 c 132 g 140 t

ORIGIN

Query Match 65.1%; Score 582.6; DB 10; Length 613;  
Best Local Similarity 99.2%; Pred. No. 4.2e-140;  
Matches 596; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 140 gcaagaaaccttccattatattcttcaagaacacttacagctgcagcagagtgatga 199  
Db 14 GCAGGAACCTTCATTAATTTCTTCAAGCACTTACAGTCTACCGAGTGGGATGA 73  
QY 200 aagttctaactcttccctcctcctcctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 259  
Db 74 AAGTTCTAATCTCTTCCCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 133  
QY 260 ggaagctgaatccaaggggtgcgcaagagcgaagagccgaagcttcaagagat 319  
Db 134 GCAGCCTGATCCAGGGGGTGCACAGAGGCGACAGGAGCCAGGACCTTTAGAGAT 193  
QY 320 ggctcagaagcgcgcaagaatgtgagtgcaaatgtgtgtgtgtgtgtgtgtgtgtgtgt 379  
Db 194 GGCTCCAGGAAGCGGCGCAAGATGTGATGAAGATTTGTTCTTGAGAGCCCGAATA 253  
QY 380 gaaatctacagcagtgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 439  
Db 254 GAAATTCATGACAGTGTCTGTGGCTGCCAAGAAACAGTGCCTGTGATCATTTCAAG 313  
QY 440 gcaatgtgaagaaacaagacacccaagagccacagaagacccaagaagcattccagag 499  
Db 314 GAATGTGAAGAAACAAGACACAAAGGACACAGAAACCAAGCATTTCCACAG 373  
QY 500 cctgcagcaatttctcaacaatgtcagctaaagaagctgtgtgtgtgtgtgtgtgtgtgtgt 559  
Db 374 CCGTCAGCAATTTCTCAACAATGTGAGTGAAGAGCTTGGCTCTTGTGAGAGGC 433  
QY 560 tctgagcgccactcttccaatlaaacattctcagccaagaagcagtgagcacactac 619







```

RESULT 6
A1809681/c 504 bp mRNA linear EST 07-JUL-1999
LOCUS w76c06.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:2361514 3', mRNA sequence.
ACCESSION A1809681
VERSION A1809681.1 GI:5396247
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 504)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Glbco
High quality sequence stop: 459.
FEATURES
source
1..504
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2361514"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pUT73D-Pac (Pharmacia) with
a modified polylinker; Site.1: Not I; Site.2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung MDHL19W, testis NRT, and B-cell
NCL1CGAP_GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
1 M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo.
BASE COUNT 134 a 85 c 141 g 144 t
ORIGIN
Query Match 54.4%; Score 487.2; DB 9; Length 504;
Best Local Similarity 99.2%; Pred. No. 2e-115;
Matches 500; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
Oy 392 cagtgctgggctgcgaagaagcagtgccctgtgatcattccaagggaatgtgaaga 451
|||||
Db 504 CAGTGTCTGGGCTG-CANAAGACAGTGCCTGTGATCATTTCAAGGGCAATGTGAAGA 446
Oy 452 aaacaagacaccaaaagcaccagaaagccaagaagcattccagagctgcagacaat 511
|||||
Db 445 AAACAGACACCAAAAGGACCAAGCAAGCAAGCAAGCATTTCCAGAGCCTGCAGCAAT 386
Oy 512 ttctcaacaatgctcagtaagaagcttgctcgtccttgtaggaagctctgcagccca 571
|||||
Db 385 TTCTCAACAATGTCAGCTAAGAACCTTGTGCTCTTTGAGGAGCTCTGAGGCGCCA 326
Oy 572 ctctccaatataacatctcagccaagaagacagtgagcacacctaccagacactctc 631
|||||
Db 325 CTCTTCCATTTAAACATTTCTCAGCCCAAGACAGACTGACACACTTACCAGACACTCTTC 266
Oy 632 ttctccacactcaactctcccaactgtaaccacccctaaatcatctcagtgcttcaaaa 691
|||||
Db 265 TTCTCCACACTCACTCTCCACTGTACCCACCCCTAAATCATTCAGAGCTCTCAAAAAG 206
Oy 692 catgttttcaagatcattgtgtgtgtcctctctcagtgcttctctctctcgtcagtc 751
|||||
Db 205 CATGTTTTCAGATCATTTTGTGTGTCCTCTCTAGTGTCTTCTCTCTCGTCACTC 146

```

```

Oy 752 tttagctgtgccctccctccctaccagcttagagcttaattaccctgaagaattccagaaa 811
|||||
Db 145 TTAGCTGTGCCCTCCCTCCCTTACCAGAGCTTAGGCTTAATTCCTGAAGATTTCCAGGAAA 86
Oy 812 ctgtagcttctcagtagtgatcatttaacctaaatgcaatcaggaagtagcaaacaga 871
|||||
Db 85 CTGAGCTTCTCAGCTAGTACTGTCATTTAATTCGAATCAGCAGCAAGTGAACAAACAGA 26
Oy 872 agtcaataataatttttaatgctc 895
|||||
Db 25 AGTCAATTAATTAATTTTAAATGTC 2
RESULT 7
A1830551/c 491 bp mRNA linear EST 21-DEC-1999
LOCUS wj51b02.x1 NCI-CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2406315 3',
DEFINITION mRNA sequence.
ACCESSION A1830551
VERSION A1830551.1 GI:5451222
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 491)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.llnl.gov/dbp/image/image.html
Insert Length: 406 Std Error: 0.00
Seq primer: -40UP from Glbco
High quality sequence stop: 463.
FEATURES
source
1..491
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2406315"
/clone_lib="NCI CGAP_Lu19"
/tissue_type="squamous cell carcinoma, poorly
differentiated (4 pooled tumors, including primary and
metastatic)"
/dev_stage="adult"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Lung; Vector: pUT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
pooled lung tumor tissue, and was then primed with a Not I
+ oligo(dT) primer. Double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pUT73 vector. Library went through one round of
normalization. Library constructed by Bento Soares and M.
Fatima Bonaldo.
BASE COUNT 129 a 81 c 132 g 145 t 4 others
ORIGIN
Query Match 52.6%; Score 471; DB 9; Length 491;
Best Local Similarity 97.1%; Pred. No. 3.1e-111;
Matches 477; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
Oy 404 tgcacaagaagcagtgccctgtgatacttcaagggaatgtgaagaacaagacacc 463

```

Db 491 TGCCAAAGACAGTGGCCCTGTGATCATTTCAAGGCGCAATGTGAGAAAACAAGACACC 432  
 Qy 464 aagagccacacagaagaacacacagatccacagagctccagcaattctcaaaaat 523  
 Db 431 AAAGGACACACAGAAACCAACAGATTCAGAGCGCTGCGACAAATTTCTCAACAAT 372  
 Qy 524 gtcagctaaagaagcttgcctcgtcttctgtagagctctgagcgccacctctccaat 583  
 Db 371 GTCAGCTAAAGAAGCTTGTCTGCTGCTTGTAGAGCTCTGAGCGCCACCTTTTCCATTA 312  
 Qy 584 aacattctcagcagaagaagcagtgagcacactaccagacactctctctccaccc 643  
 Db 311 AACCTTCMCAGCCAAAGAGAGAGAGACANNACAGANCTCTCTCTCCACCTC 252  
 Qy 644 actctccacatgtacacaccccaatcatctccagtgctctcgaagaagcattttc 703  
 Db 251 ACCTCTCCACATGTACACACCCCTAAATCATTCAGTCTCTCAAAAAGCATGTTTTC 192  
 Qy 704 gatacattgtctgtgtctctctcagtgctctctctctcgtcagctcagctgtgc 763  
 Db 191 GATCATTTTGTGTTGCTCTCTCTAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 132  
 Qy 764 ctccctaccacagctcagctcagctcagctcagctcagctcagctcagctcagct 823  
 Db 131 CTCCTCTTACCCAGGCTTAAAGCTTAACTTACCTGAAAGATTCAGGAACTGTAGCTTCT 72  
 Qy 824 agctagtgatcattcaatcaatgcaatcaggaagaagtagcaacaagaatcaata 883  
 Db 71 AGCTAGTGATTTTAACTTAATGCAATGAGAAAGTACGAAACAGAAAGTCAATTAATA 12  
 Qy 884 ttcttaaatgt 894  
 Db 11 TTTTAAATGT 1

RESULT 8  
 BP061539 482 bp mRNA linear EST 16-OCT-2000  
 LOCUS 7152910.x1 Soares\_NSF\_F8\_9W\_OT\_PA\_P\_S1 Homo sapiens cDNA clone  
 DEFINITION IMAGE:3390114 3', mRNA sequence.  
 ACCESSION BP061539  
 VERSION BP061539.1 GI:10820370  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 482)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -400p from Gibco  
 High quality sequence stop: 470.  
 Location/Qualifiers  
 1..482  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="3390114"  
 /clone\_lib="Soares\_NSF\_F8\_9W\_OT\_PA\_P\_S1"  
 /lab\_host="DH10B"  
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with  
 a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
 Equal amounts of plasmid DNA from five normalized  
 libraries were mixed, and ss circles were made in vitro.  
 Following HAP purification, this DNA was used as tracer in  
 a subtractive hybridization reaction. The driver was  
 PCR-amplified cDNAs from pools of 5,000 clones made from

the same 5 libraries. The pools consisted of the following  
 libraries and clones: Soares NB2HP pool 1:  
 309384-310919, 322308-325895 Soares NB2HP pool 1:  
 145032-147335, 147720-148103, 148872-149255, 15002 -  
 150407, 151176-152327 Soares NB2HP-9W pool 1:  
 758280-760583, 772104-774407 Soares NBHPA pool 1:  
 304776-306311, 320136-322823, 326280-326663 Soares NBHP  
 pool 1: 723720-726407, 739080-740999 Subtraction by Bento  
 Soares and M. Fatima Bonaldo."

BASE COUNT 131 a 77 c 134 g 140 t  
 ORIGIN

Query Match 52.0%; Score 465.2; DB 10; Length 482;  
 Best Local Similarity 99.2%; Pred No. 9.7e-110;  
 Matches 478; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Qy 415 cagtgccctgtgatacttcaaggagcaatgtgaagaacaacagacacagcacc 474  
 Db 482 CAGTGCCCTGTGATCATTTCAAGGCGCAATGTGAGAAAACAAGACACCAAGGACGCAC 423  
 Qy 475 agaaagccaaacagaagatccagaagctgcagcaaatctcaacaatgtcagtaaga 534  
 Db 422 AGAAAGCCAAACAGCATTCAGAGCCTGCAGCAATTTCTCAACCAATGTACAGTAAGA 363  
 Qy 535 agcttgctcgtcctgtgtagagagctcagagccacatctcaatlaaacatctcag 594  
 Db 362 AGCTTGCTCTGCTCTTGTAGAGAGCTCTAGCCGCCACCTCTTCATTTAAACATTTCC 303  
 Qy 595 ccaagaagaagtgagcacact-accagacactctctctccacactctccac 653  
 Db 302 CCAAGAAGACAGTGAAGCACACCTTAACAGACACTTTTCTTTCCACCTCACTGCCAC 243  
 Qy 654 tgtaccaccccaatactctcagtgctcctcaaaaagatgttttcaagaatcttg 713  
 Db 242 TGTACCCACCCCTAAATCTTCAGTGTCTCAAAAAGCATTTTTCAGATCAATTTTG 183  
 Qy 714 ttgtgtcctctcagtgctctctctctcctcgtcagctcagctgcccctccatc 773  
 Db 182 TTTGTTGCTCTCTAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 123  
 Qy 774 ccagcctagagcttaattacactgaagaagatccagaagaactgtagctctcagtagtgc 833  
 Db 122 CCAGGCTTAGGCTTAATCTCTGAAGAAGATTCAGAAACTGTAGCTCTACTAGTGTGC 63  
 Qy 834 attaaccttaatacgaatcagaagaagtagcaagaagaagtcataatattttaagt 893  
 Db 62 ATTTAACCTTAATGCAATCAGGAAGTAGCAACAGAGATCAATTAATTTTAAATG 3  
 Qy 894 tc 895  
 Db 2 TC 1

RESULT 9  
 A1142875/c 477 bp mRNA linear EST 23-OCT-1998  
 LOCUS A1142875  
 DEFINITION 0x64e06.s1 Soares\_NHMPu\_S1 Homo sapiens cDNA clone IMAGE:1661122  
 3', mRNA sequence.  
 ACCESSION A1142875  
 VERSION A1142875.1 GI:3659234  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 477)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov

This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 889 Std Error: 0.00  
 Seq primer: -40ml3 fwd. ER from Amersham  
 High quality sequence stop: 383.

## FEATURES

## source

Location/Qualifiers

1. .477  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1661122"  
 /clone\_lib="Soares\_NHMPu\_S1"  
 /tissue\_type="Pooled human melanocyte, fetal heart, and pregnant uterus"  
 /lab\_host="DH10B"

/note="Organ: mixed (see below); Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NbHM, pregnant uterus NbHPU, and fetal heart NbH19W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."

## BASE COUNT

129 a 74 c 135 g 139 t

## ORIGIN

## Query Match

Best Local Similarity 99.8%; Score 464; DB 9; Length 477;  
 Matches 475; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 420 cccctgtatattcagaaggagtgagaagaacaccaccagaagcaccaccagaaga 479  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 477 CCCGTGATCATTTTCAAGGGCAATGTGAAGAAACAAGACACCAAGCACCACAGAAA 418  
 OY 480 gccacaagaatcaccagaagcctgaccagaattctcacaacatgtaagaagctt 539  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 417 GCCAACAAGCATTTCCAGAGCGCTGCCAGCAATTTCTCAACAATGTCAAGAAAGCTT 358  
 OY 540 tgcctgcctttgttagagagctctgagcgccacatctccaatlaaacattcagccaag 599  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 357 T-CTTGCTTGTGTAGGAGCTGTGAGCGCCCACTCTTCCAAATTAACAATTTCAGCAGAA 299  
 OY 600 aagaagagagacacccatcaccagaacatcttcttccacatcactctccacatgtaac 659  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 298 AAGACAGTGAACACACCTACGACACTCTTCTCCACCTCACTCCACACTGTACC 239  
 OY 660 caccctaatcattccagtgctcacaagaacatglttttaagatcatattgtttgt 719  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 238 CACCCCTAATATCATTCAGTGTCTTCAAAAGCATGTTTTCAGATCATTTTGTGTTT 179  
 OY 720 gctctctagtgcttcttctctgcagctttagcctgtgacctccctaccagaagc 779  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 178 GCTCTCTAGTGTCTTCTCTCTGCTAGCTTGTAGCCGTGCTCCCTTACCAGAGC 119  
 OY 780 ttagcttaattacatggaagatccaggaacatgtagttctctagctagtgatcattaa 839  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 118 TTAGCTTAATTAATCTGAAAGATTCAGAGAACTGTAGCTTCTGTAGCTCATTTAA 59  
 OY 840 ccttaaatgcaatcaggaagtagcaacaagaatcaataatattttaaatgtc 895  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 58 CCTTAAATGCAATCAGAAAGTAGCAACAGAAATCAATTAATTTTAAATGTC 3

## RESULT 10

AI276107/c

## LOCUS

AI276107 461 bp mRNA linear EST 29-JAN-1999

## DEFINITION

q174h09.x1 Soares\_NHMPu\_S1 Homo sapiens cDNA clone IMAGE:1878113

## ACCESSION

AI276107

## VERSION

AI276107.1 GI:3898381

## KEYWORDS

EST.

## SOURCE

human.

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

## REFERENCE

1 (bases 1 to 461)

## AUTHORS

NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.

## TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

## JOURNAL

Tumor Gene Index

## COMMENT

Unpublished (1997)

## CONTACT

Contact: Robert Strausberg, Ph.D.

## EMAIL

Email: c9apbs-r@emall.nih.gov

## TEXT

This clone is available royalty-free through LNL ; contact the

## IMAGE

IMAGE Consortium (info@image.llnl.gov) for further information.

## INSERT

Insert Length: 412 Std Error: 0.00

## Seq

Seq primer: -40UP from Glibco.

## FEATURES

## source

Location/Qualifiers

1. .461  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1878113"  
 /clone\_lib="Soares\_NHMPu\_S1"  
 /tissue\_type="Pooled human melanocyte, fetal heart, and pregnant uterus"  
 /lab\_host="DH10B"

/note="Organ: mixed (see below); Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NbHM, pregnant uterus NbHPU, and fetal heart NbH19W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."

## BASE COUNT

122 a 74 c 128 g 136 t

## ORIGIN

## Query Match

Best Local Similarity 98.9%; Score 431.8; DB 9; Length 461;  
 Matches 455; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

OY 436 aagggcaatgtgaagaaacacacacacacacacacacacacacacacacacacac 495  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 461 AAGGGCAATGTGAAGAAACAAGACCCCAAGGGCACACAGAAAGCAAGACATTC 402  
 OY 496 agagcctgccaagaattctcacaacatgtaagaagctttgctgctgttag 555  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 401 AAGGCCNGCCAGCAATTTCTCAACAATGTCAAGCTAAGAAAGCTTGTGCTGTGAG 342  
 OY 556 gagctctgagggccacatcttccaatlaaacattctcagccaagaagacagtgaacac 615  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 341 GAGCTCTGAGCCGCCCACTTCCAAATTAACAATTCACGCCAAGAAAGACAGTGAAC 281  
 OY 616 ctaccagaacatcttcttccacatcacttccacatgtaacacacacacacacacac 675  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 282 CTACCAAGACAGT-TTCTTCCCACTCACTCTCCACAGTACCCACCCCTTAATATTC 224  
 OY 676 cagtgctcctcaaaagacatgttttccaagatcatttgtttgtgctctctagtgct 735  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 223 CAGTGTCTCAAAAAGCATGTTTTCAGATCATTTTGTGTTGTGCTCTCTAGTGTCT 164  
 OY 736 tcttctctgtaagcttagcctgtgacctccctaccacagagcttagcctaataact 795  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 163 TCTTCTCTGTAAGTCTTGTAGCTGTGCTCCCTCCCTTACCAGCTTAAATTAATCC 104  
 OY 796 gaaagattccagaagacatgtaagcttctctagctagtgtaatttaaccttaagcaatg 855  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 103 GAAAGATTCACAGAAACTGTACCTTCTGTAGCTGTAGTGTCTTAATCCCTTAATGCAATGAC 44  
 OY 856 gaaagtagcaacaacagagcacaataataattttaatgctc 895

| Db                        | 43   | GAAGTAGCAACAGAGTCATTAATTTTAAATGTC                              | 4                 |
|---------------------------|--|--|-------------------|
| RESULT                    | 11   |  |                   |
| LOCUS                     | BE148882   | 563 bp   | mRNA linear       |
| DEFINITION                | CM2-HT0244-221199-038-f02 HT0244   | Hom sapiens cDNA,  | mRNA sequence.    |
| ACCESSION                 | BE148882   |  |                   |
| VERSION                   | BE148882.1   | GI:8611606   |                   |
| KEYWORDS                  | EST.   |  |                   |
| SOURCE                    | human.   |  |                   |
| ORGANISM                  | Hom sapiens  |  |                   |
| REFERENCE                 | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.   |  |                   |
| AUTHORS                   | 1 (bases 1 to 563)<br>Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Brites, M.R., Nagai, M.A., da Silva, W. J., Zago, M.A., Bordin, S., Costa, F.F., Godman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A.J., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.   |  |                   |
| TITLE                     | Shotgun sequencing of the human transcriptome with ORF expressed sequence tags   |  |                   |
| JOURNAL                   | Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)   |  |                   |
| MEDLINE                   | 20202663   |  |                   |
| COMMENT                   | Contact: Simpson A.J.G.<br>Laboratory of Cancer Genetics<br>Ludwig Institute for Cancer Research<br>Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil<br>Tel: +55-11-2704922<br>Fax: +55-11-2707001<br>Email: asimpson@ludwig.org.br<br>This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be found in the following URL<br>(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-CM2-HT0244-221199-038-f02&t3=1999-11-22&t4=1)<br>Seq primer: puc 18 forward<br>High quality sequence start: 26<br>High quality sequence stop: 562.<br>Location/Qualifiers<br>1..563<br>/organism="Homo sapiens"<br>/db_xref="taxon:9606"<br>/clone_lib="HT0244"<br>/dev_stage="Adult"<br>/note="Organ: head,neck; Vector: puc18; Site.1: Smal; Site.2: Smal; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - ludwig institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions." |  |                   |
| FEATURES                  |  |  |                   |
| SOURCE                    |  |  |                   |
| BASE COUNT                | 159 a  | 161 c  | 153 t             |
| ORIGIN                    |  |  |                   |
| Query Match               | 47.8%  | Score 427.6;   | DB 9; Length 563; |
| Best Local Similarity     | 98.9%  | Pred. No. 5,4e-100;  |                   |
| Matches 441; Conservative | 0;   | Mismatches 4;  | Indels 1; Gaps 1; |
| QY                        | 450  | gaataacagacacccaagaagcaccagaagaacccaagaagattccagagctgcccagca   | 509               |
| Db                        | 22   | GAATTCAGGACACCAAGAGCA-CACAGAAAGCCAAAGCATTCACAGAGCTCCAGCA       | 80                |
| QY                        | 510  | attctccaacaatgctcagctaaagaagcttgctgctgctgcttgtagagctctgagcgc   | 569               |
| Db                        | 81   | ATCTCTCAAAACATGTGAGTAAAGAGCTTGCTGCTCTTTAGAGAGCTCTAGAGGCC       | 140               |
| QY                        | 570  | cacctcttcacattaaattctcagccaagaagacagtgtgacacactaccagaactct     | 629               |
| Db                        | 141  | CACCTCTTCACATTAACCATTTCTCAGCCCAAGAAAGCACTGTGACACACCTACGACACTCT | 200               |

|    |     |  |     |
|----|-----|--|-----|
| QY | 630 | tctctcccaacctactctcccaactctaccaccccttaattccagtgcttcaaaa          | 689 |
| Db | 201 | tctctcccaacctactctcccaactctaccaccccttaattccagtgcttcaaaa          | 260 |
| QY | 690 | agcatgtttttcaagaatcatcttgtttgctctctcctagtctctctctctgcag          | 749 |
| Db | 261 | agcatgtttttcaagaatcatcttgtttgctctctcctagtctctctctctgcag          | 320 |
| QY | 750 | tcttagccctgtccctccccccttaccacgaagcttaagcttaattaccctgaagaattccaga | 809 |
| Db | 321 | tcttagccctgtccctccccccttaccacgaagcttaagcttaattaccctgaagaattccaga | 380 |
| QY | 810 | aactctagcttctcctagctagtgtgcatltaacctaaatgcaatcgaaagatgacaaaca    | 869 |
| Db | 381 | aactctagcttctcctagctagtgtgcatltaacctaaatgcaatcgaaagatgacaaaca    | 440 |
| QY | 870 | gaagtaacaataaatttttaaatgtc                                       | 895 |
| Db | 441 | gaagtaacaataaatttttaaatgtc                                       | 466 |

|            |  |        |       |        |                 |  |
|------------|--|--------|-------|--------|-----------------|--|
|            | RESULT   | 12     |       |        |                 |  |
|            | AA479746/c   |        |       |        |                 |  |
| LOCUS      | AA479746   | 426 bp | mRNA  | linear | EST 08-AUG-1997 |  |
| DEFINITION | z35g03.s1 Soares ovary tumor NBH0T Homo sapiens cDNA clone   |        |       |        |                 |  |
| ACCESSION  | IMAGE:740020 3'  |        |       |        |                 |  |
| VERSION    | AA479746   |        |       |        |                 |  |
| KEYWORDS   | AA479746.1 GI:2205632  |        |       |        |                 |  |
| SOURCE     | Est.   |        |       |        |                 |  |
| ORGANISM   | human.   |        |       |        |                 |  |
| REFERENCE  | Homo sapiens   |        |       |        |                 |  |
| AUTHORS    | Eukariyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;<br>Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.<br>1 (bases 1 to 426)<br>Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,<br>Kucaba,T., Lacy,M., Le,N., Lennon,G., Harris,M., Martin,T., Moore,B.,<br>' T., Waterston,R., Steptoe,M., Tan,F., Theising,B., White,Y., Wyllie,<br>Schellenberg,K., R., Wilson,R.<br>WashU-Merck EST Project 1997<br>Unpublished (1997)                           |        |       |        |                 |  |
| TITLE      | Contact: Wilson RK   |        |       |        |                 |  |
| JOURNAL    | Washington University School of Medicine   |        |       |        |                 |  |
| COMMENT    | 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108<br>Tel.: 314 286 1800<br>Fax: 314 286 1810<br>Email: east@watson.wustl.edu<br>This clone is available royalty-free through LLNT ; contact the<br>IMAGE Consortium ( <a href="http://infoimage.llnl.gov">infoimage.llnl.gov</a> ) for further information.<br>Seq primer: -41m13 fwd_ RT from Amersham<br>High quality sequence stop: 419.   |        |       |        |                 |  |
| FEATURES   | Location/Oualifiers  |        |       |        |                 |  |
| Source     | 1..426   |        |       |        |                 |  |
|            | /organism="Homo_sapiens"   |        |       |        |                 |  |
|            | /db_xref="GDB:5939727"   |        |       |        |                 |  |
|            | /db_xref="taxon:9606"  |        |       |        |                 |  |
|            | /clone="IMAGE:740020"  |        |       |        |                 |  |
|            | /clone_id="Soares ovary tumor NBH0T"   |        |       |        |                 |  |
|            | /sex="Female"  |        |       |        |                 |  |
|            | /tissue_type="ovarian tumor"   |        |       |        |                 |  |
|            | /lab_host="DH10B (ampicillin resistant)"   |        |       |        |                 |  |
|            | /note="Organ: ovary; Vector: p773D (Pharmacia) with a<br>modified polylinker. Site.1: Not I; Site.2: Eco RI; 1st<br>strand cDNA was primed with a CCGTTCATCTTTTCTTTTTT 3']<br>TGTTACCAGCATCAGAGGGAGCGGCCGTTCATCTTTTCTTTTTT 3']<br>double-stranded cDNA was size selected, ligated to Eco RI<br>adapters (Pharmacia), digested with Not I and cloned into<br>the Not I and Eco RI sites of a modified p773 vector<br>(Pharmacia). Library constructed by Bento Soares and<br>M.Fatima Bonaldo." |        |       |        |                 |  |
| BASE COUNT | 121 a  | 64 c   | 122 g | 119 t  |                 |  |
| BRIGIN     |  |        |       |        |                 |  |

Query Match 47.6%; Score 426; DB 9; Length 426;  
Best Local Similarity 100.0%; Pred. No. 1.3e-99;  
Matches 426; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 469 caccacagaagcacaagaacatccagagcctgcagcaatcttccaacatgctcag 528  
DB 426 CACCACAGAAAGCAACAAACATTCAGAGCCTGCCAGCAATTTCTCAAAACATGTGAG 367  
QY 529 ctaagaagcttgctgccttgtagagctctgagagcccaactctccaataaacat 588  
DB 366 CTAAGAAGCTTGGCTGCTGCTTTGTAGAGCTGTGAGGCCACACTTTCATTAATTAACAT 307  
QY 589 tctacagccaagaagacagctgagcaccctacagacaactctctctcccaactcact 648  
DB 306 TCTCAGCCAAAGACAGACAGTGCACACCTACAGACACTCTTCTCCACCTACACTCT 247  
QY 649 cccactgtaccacccccaataatcaltccagtgctctcaaaaagaatgtttccaagatca 708  
DB 246 CCCACTGTACCACCCCTTAATCAATTCAGTCTCTCAAAAAGCATGTTTTCAGATCA 187  
QY 709 ttgtgtgtgtcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 768  
DB 186 TTGTGTGTGTGCTCTCTCTAGTGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 127  
QY 769 ctacccagcttagagcttaattactcctgaagaatccaggaactgtagctctcctagcta 828  
DB 126 CTTACCCAGGCTTAGGCTTAATTAATCTGAAAGATTCCAGGAACGTGAGCTTCCAGCTA 67  
QY 829 gtgtcatttaaccttaaatgcaatcagaagaatgagcaagaatgcaataataatttt 888  
DB 66 GTGTCAATTAACCTTAATGCAATCAGGAAGTAGCAAGCAAGCAATTAATTAATTTT 7  
QY 889 aaatgt 894  
DB 6 AAATGT 1

RESULT 13  
AA430487/c 419 bp mRNA linear EST 16-OCT-1997  
LOCUS  
DEFINITION  
IMAGE:770147.3', mRNA sequence.  
ACCESSION  
AA430487  
VERSION  
AA430487.1 GI:2111061  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 419)  
AUTHORS  
Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,  
Kucab,T., Lacy,M., Le,N., Lennon,G., Maira,M., Martin,J., Moore,B.,  
Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wyllie,  
'T', Waterston,R. and Wilson,R.  
Washu-Merck EST Project 1997  
Unpublished (1997)  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu

FEATURES  
SOURCE  
1. .419  
/organism="Homo sapiens"  
/db\_xref="GDB:596002"  
/db\_xref="taxon:9606"  
/clone="IMAGE:770147"

/clone\_1lb="Soares ovary tumor NbHOT"  
/sex="Female"  
/tissue\_type="Ovarian tumor"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Ovary; Vector: p773D (Pharmacia) with a  
modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTACCAATCTGAAGTGGAGGCGCGGGGTTTTTTTTTTTTTTT 3'],  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified p773 vector  
(Pharmacia). Library constructed by Bento Soares and  
M.Fatima Bonaudo."

Query Match 46.3%; Score 414.8; DB 9; Length 419;  
Best Local Similarity 99.5%; Pred. No. 1.1e-96;  
Matches 416; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 478 aagccaagaagcattccagagcctgcagcaatcttccaacatgctcagtaagaagc 537  
DB 419 AAGCCAAACAAACATTCAGAGCCTGCAGCAATTTCTCAAAACATGTGAGTAAAGC 360  
QY 538 ttgtcctgccttgtagagctctgagagcccaactctccaataaacatccagcca 597  
DB 359 TTGTGCTGTGCTTTTAGAGAGCTGTGAGGCCACCTATTTCCATTAATTAATTTCTAGCCA 300  
QY 598 agaagacagtagcacacactcaccagacactctctctcccaactcactccactccta 657  
DB 299 AGAAGACAGTGAAGCAGACCTACCAACACTCTTCTCTCCACCTCAGCTCCACTGTA 240  
QY 658 cccacccccaataatcattccagtgctctcaaaaagaatgtttccaagatcattgtttg 717  
DB 239 CCCACCCCTTAATCAATTCAGAGCTCTCAAAAAGCATGTTTTCAGATCAATTTTGTG 180  
QY 718 ttgtcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 777  
DB 179 TTGCTCTCTTAGTGTCTTCT 120  
QY 778 gcttagcttaattactcctgaagaatccaggaactgtagctcctcctagctagctcatt 837  
DB 119 GCTTAGGCTTAATTAATTCGAAAGATTCCAGGAACGTGAGCTTCCAGCTAGTATGTGATTT 60  
QY 838 aaccttaaatgcaatcagaagaatgagcaagaatgcaataataataatgtc 895  
DB 59 AACCTTAATGCAATCAGGAAGTAGCAAGCAAGATCAATTAATTTTAAATGTC 2

RESULT 14  
AM298444/c 429 bp mRNA linear EST 16-JAN-2000  
LOCUS  
DEFINITION  
IMAGE:2732243.3', mRNA sequence.  
ACCESSION  
AM298444  
VERSION  
AM298444.1 GI:6705080  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 429)  
AUTHORS  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaps-r@mail.nih.gov

FEATURES  
SOURCE  
1. .419  
/organism="Homo sapiens"  
/db\_xref="GDB:596002"  
/db\_xref="taxon:9606"  
/clone="IMAGE:770147"

NCI-CGAP clone distribution information can be found through the  
I.M.A.G.E. Consortium/ILNL at:  
www.bio.lnl.gov/birop/image/image.html  
Seq primer: M13 Forward  
POLY-A-yes

## FEATURES

Source

Location/Qualifiers

1. 429  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2732243"  
/clone\_lib="NCI-CGAP\_Sub6"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pTR73D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; NCI CGAP Sub6  
is a subtracted library derived from BW, which consists of  
a mixture of four normalized libraries: NCI CGAP\_Brn50,  
NCI CGAP\_Lu13, NCI CGAP\_Ov18, GBC1. The NCI CGAP Sub6  
library had 7 million recombinants. A single-stranded DNA  
preparation of BW was used as a tracer in a subtractive  
hybridization with a driver comprising the IMAGE pool  
(NCI CGAP\_Kid3 pool 1, LAM 3334-3337, 3682-3683,  
3798-3803 (IMAGE Clonoids 1322376-1323911,  
1456008-1456775, 1500552-1502855); NCI CGAP\_Kid5 pool 1  
LAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clonoids  
1323912-1325831, 1471368-1472903, 1492104-1493255);  
NCI CGAP\_Lu5 pool 1 LAM 3575-3582, 3851-3854 (IMAGE  
Clonoids 1414920-1417991, 1520904-1522439); NCI CGAP\_GC4  
pool 1 LAM 3164-3167, 3716-3720, 3733-3735 (IMAGE  
Clonoids 1257096-1258631, 1469064-1470983, 1475592-1476743  
); NCI CGAP\_Pr22 pool 1 LAM 2457-2459, 2758-2759,  
3062-3068 (IMAGE Clonoids 985608-986759, 1101192-1101959,  
1217928-1220615); NCI CGAP\_Co10 pool 1 LAM 2644-2653,  
2871-2872 (IMAGE Clonoids 1057416-1061255, 1144584-1145351  
). (50% of the driver population), plus a pool of 3,840  
arrayed clones from NCI CGAP\_Sub1 (IMAGE Clonoids  
2708616-2710535) and NCI CGAP\_Sub2 (IMAGE Clonoids  
2710536-2712455) (20% of the driver population), plus a  
pool of 11,136 clones from NCI CGAP\_Sub3 (IMAGE Clonoids  
2712456-2723591) (30% of the driver population).  
Subtraction was performed as previously described (Bonaldi  
, Lennon & Soares (1996): Normalization and Subtraction:  
Two Approaches To Facilitate Gene Discovery. Genome  
Research 6, 791-806  
TAG\_LIB=NCI CGAP\_Lu13  
TAG\_TISSUE=Lung  
TAG\_SEQ=GCCGG"

BASE COUNT 92 a 82 c 112 g 140 t 3 others  
ORIGIN

Query Match 45.7%; Score 409; DB 9; Length 429;  
Best Local Similarity 99.3%; Pred. No. 3,4e-95;  
Matches 409; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

312 tagagatgtcccaaggaagcgcccaagaatgtatgtaagttctctagagc 371  
DB 429 TAGGAGATGCTCCAGGAAGGCGGCCAGAAATGTGATCAAGAGATGGTCTGAGGC 370  
372 cccgaagaagaatcattacacagtgctgtggtccaaagaagcgctgtgatatca 431  
DB 369 CCCGAGAAAGAAATTCATGACAGTGTCTGGCTGCCAAGAAGCAGTGCCTGTGATCA 310  
399 TTTCAAGGGCATGTGAAGAAAAGACACCAAGGCGCCACAGAAAGCCAAACAAGCA 250  
432 ttcaagggcaatgtgaagaagaacagacacaggaagccacagagaagccaaacaagca 491  
DB 309 TTTCAAGGGCATGTGAAGAAAAGACACCAAGGCGCCACAGAAAGCCAAACAAGCA 250  
492 ttcaagggcctccaggaagcatttctaaacaatgtcagctaaagaagcttgcctctt 551  
DB 249 TTCCAGAGCTGCGCAGCAATTTCTCAACAAATGTCAGCTTAAGAAAGCTTGGCTCTCCCTTT 190  
552 gtaagagctctgaagcccaactctcaatlaaacattctcagcagaagaagaagcagtgagc 611  
DB 189 GTAGAGGCTGTGAGGCGCCACACTTTCATTTAAACATTTCTCAGCCAAAGAGACAGTGAAC 130

QY 612 acacctaccagaacacctcttctccacactcactccactgtatccacccttaac 671  
DB 129 ACACCTACGAGACACCTCTTCTCCACCTCCTCCACCTCCACAGTACCCACCTTAATC 70  
QY 672 attcagtgctctcaaaagacatgtcttcaagatcatttgttctgtc 723  
DB 69 ATTCCAGTGTCTCTCAAAAGCATGTTTTCAGATCATTTTGTGTGCTC 18

RESULT 15  
A1079465/C 431 bp mRNA linear EST 01-OCT-1998  
LOCUS 0239f05.s1 Soares.NHMPu\_S1 Homo sapiens cDNA clone IMAGE:1677729  
DEFINITION 3', mRNA sequence.  
ACCESSION A1079465  
KEYWORDS A1079465.1 GI:3415716  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 431)  
NCI CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-remail.nlm.nih.gov  
This clone is available royalty-free through ILNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Insert Length: 921 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 431.  
Location/Qualifiers

## FEATURES

Source

1. 431  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1677729"  
/clone\_lib="Soares\_NHMPu\_S1"  
/tissue\_type="Pooled human melanocyte, fetal heart, and  
pregnant uterus"  
/lab\_host="DH10B"  
/note="Organ: mixed (see below); Vector: pTR73D-Pac  
(Pharmacia) with a modified polylinker; Site 1: Not I;  
Site 2: Eco RI; Equal amounts of plasmid DNA from three  
normalized libraries (melanocyte 2NbHM, pregnant uterus  
NBHPU, and fetal heart NBH19W) were mixed, and ss circles  
were used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from pools of  
5,000 clones made from the same 3 libraries. The pools  
consisted of I.M.A.G.E. clones 260232-265223,  
340488-345479, and 484488-489479."

BASE COUNT 121 a 65 c 124 g 121 t  
ORIGIN

Query Match 45.3%; Score 405; DB 9; Length 431;  
Best Local Similarity 99.5%; Pred. No. 3.7e-94;  
Matches 427; Conservative 0; Mismatches 0; Indels 2; Gaps 2;  
468 goaccacagaagccaacaagcattccagagcctgcagcaatttccaacaatgtca 527  
DB 431 GCACCAGAGAAAGCAACAGCATTCAGAG-CTGCCAGCAATTTCTCAACAATGTCA 373  
558 gctaaagacgttctgtccttctgtagg-agctctgagcgcccaacttccaatlaaac 586  
DB 372 GCTAAGAGCTTTGCTGTGCTTTGTAGCAGCTCTGAGGCCCCACCTTCCATTAAC 313  
587 attctcagcagaagaagatgagacacactaccagacactcttctccactcaact 646  
DB 312 ATTTCAGCCAAAGAGACAGTGAAGACACTTACAGACAGACACTTCTCTCCACCTCACT 253

```

Qy 647 ctcccactgtaccacccctaaatcattccagtgctctcaaaaagcatgtttttcaagat 706
    |||
Db 252 ctcccactgtaccacccctaaatcattccagtgctctcaaaaagcatgtttttcaagat 193
    |||
Qy 707 caattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 766
    |||
Db 192 CATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 133
    |||
Qy 767 cccttaaccaggttaggttaattacccctgaagaattccaggaactgttagcttcttagc 826
    |||
Db 132 CCTTACCCAGGCTTAGGCTTAATTACCTGAAGATTCCAGGAACGTGAGCTTCCTAGC 73
    |||
Qy 827 tagtgatcattaaacctaaatgcgaatcagaagaagtagcaacagaagtcataaatattt 886
    |||
Db 72 TAGTGTCATTTAACCTTAATGCAATCAGSAAAGTAGCAACAGAAAGTCAATTAATATT 13
    |||
Qy 887 ttaaatgtc 895
    |||
Db 12 TTAATGTC 4
    |||

```

Search completed: September 16, 2002, 01:42:54  
 Job time: 3307 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 03:26:45 : Search time 89 Seconds  
(without alignments)  
2470.135 Million cell updates/sec

Title: US-09-700-770-5

Perfect score: 895  
Sequence: 1 ctaatctgtacgtacagc.....aataaatttttaaatgtc 895

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Database : Issued\_Patents\_NA.\*

1: /cgn2\_6/ptodata/2/1na/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/1na/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/1na/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/1na/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/1na/PTC05\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/1na/Backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | % Match | Query Length | ID | Description       |
|------------|-------|---------|--------------|----|-------------------|
| 1          | 56.8  | 6.3     | 7218         | 1  | US-08-232-463-14  |
| 2          | 38.6  | 4.3     | 7218         | 1  | US-08-232-463-14  |
| 3          | 38.6  | 3.8     | 289          | 4  | US-09-007-005-17  |
| 4          | 33.8  | 3.8     | 289          | 4  | US-09-244-796-17  |
| 5          | 32.8  | 3.7     | 445          | 4  | US-09-247-155-53  |
| 6          | 32.8  | 3.7     | 2686         | 4  | US-09-228-986-3   |
| 7          | 32.8  | 3.7     | 12001        | 1  | US-08-458-568A-11 |
| 8          | 32.2  | 3.6     | 6755         | 3  | US-08-931-999-4   |
| 9          | 31.6  | 3.5     | 1146         | 4  | US-08-596-684F-1  |
| 10         | 31.2  | 3.5     | 2264         | 4  | US-09-126-109-9   |
| 11         | 30.8  | 3.4     | 1698         | 1  | US-08-592-126-62  |
| 12         | 30.8  | 3.4     | 2062         | 1  | US-08-073-383-5   |
| 13         | 30.8  | 3.4     | 2062         | 3  | US-08-328-238A-3  |
| 14         | 30.8  | 3.4     | 2062         | 5  | PCT-US94-06365-5  |
| 15         | 30.8  | 3.4     | 2062         | 5  | PCT-US95-13661-3  |
| 16         | 30.8  | 3.4     | 13011        | 2  | US-08-791-849A-14 |
| 17         | 30.6  | 3.4     | 1311         | 4  | US-08-584-760A-66 |
| 18         | 30.6  | 3.4     | 7379         | 4  | US-09-341-587-5   |
| 19         | 30.4  | 3.4     | 441529       | 4  | US-09-103-840A-1  |
| 20         | 30.2  | 3.4     | 7475         | 2  | US-08-971-036-1   |
| 21         | 30.2  | 3.4     | 7475         | 3  | US-09-096-570-1   |
| 22         | 30    | 3.4     | 277          | 4  | US-08-905-223-174 |
| 23         | 30    | 3.4     | 2218         | 2  | US-08-421-044-1   |
| 24         | 29.8  | 3.3     | 2606         | 4  | US-09-252-816A-7  |
| 25         | 29.8  | 3.3     | 3439         | 4  | US-09-252-816A-6  |
| 26         | 29.8  | 3.3     | 4168         | 3  | US-08-836-567-11  |
| 27         | 29.6  | 3.3     | 1100         | 2  | US-08-481-814A-5  |

|      |      |     |      |   |                   |
|------|------|-----|------|---|-------------------|
| C 28 | 29.6 | 3.3 | 2323 | 4 | US-08-368-776A-6  |
| C 29 | 29.6 | 3.3 | 2323 | 5 | PCT-US96-00419-6  |
| C 30 | 29.6 | 3.3 | 2338 | 1 | US-08-425-069-1   |
| C 31 | 29.6 | 3.3 | 2338 | 2 | US-08-317-844B-1  |
| C 32 | 29.6 | 3.3 | 2456 | 1 | US-07-882-711-1   |
| C 33 | 29.6 | 3.3 | 2456 | 2 | US-08-462-174-1   |
| C 34 | 29.6 | 3.3 | 2457 | 2 | US-08-723-415B-12 |
| C 35 | 29.6 | 3.3 | 2457 | 3 | US-09-189-627A-12 |
| C 36 | 29.6 | 3.3 | 2492 | 1 | US-08-139-937-13  |
| C 37 | 29.6 | 3.3 | 2492 | 5 | PCT-US93-11310-13 |
| C 38 | 29.6 | 3.3 | 2517 | 1 | US-08-306-691B-18 |
| C 39 | 29.6 | 3.3 | 2517 | 2 | US-08-385-142-2   |
| C 40 | 29.6 | 3.3 | 2517 | 2 | US-08-481-814A-1  |
| C 41 | 29.6 | 3.3 | 2517 | 2 | US-08-801-092-2   |
| C 42 | 29.6 | 3.3 | 2517 | 4 | US-09-517-584A-3  |
| C 43 | 29.6 | 3.3 | 2901 | 4 | US-08-368-776A-4  |
| C 44 | 29.6 | 3.3 | 2901 | 5 | PCT-US96-00419-4  |
| C 45 | 29.6 | 3.3 | 4304 | 4 | US-08-368-776A-1  |

#### ALIGNMENTS

RESULT 1  
US-08-232-463-14  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: PTZgpt-F1s  
; US-08-232-463-14

Sequence 6, Appl  
Sequence 6, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 12, Appl  
Sequence 12, Appl  
Sequence 13, Appl  
Sequence 13, Appl  
Sequence 18, Appl  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 3, Appl  
Sequence 4, Appl  
Sequence 4, Appl  
Sequence 1, Appl

```

Query Match                               6.3%, Score 56.8; DB 1, Length 7218;
Best Local Similarity 8.8%, Pred. No. 3.9e-08;
Matches 28; Conservative 170; Mismatches 122; Indels 0; Gaps 0;

QY 537 ctttgcctcgccttgttagagagctcgtagcgcgcacactctccataatcaacttcacgc 596
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db 1171 yyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyy 1230

QY 597 aagaaagacagttagcacacctaccagaacactctctctccacactcaactccacagt 656
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1231 yyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyy 1290

QY 657 accgacccctaaatcatctccagtcgtctccaaagcagtttttcaagaatcttgc 716
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1291 yyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyy 1350

QY 717 gttgcctctcctagtcctctctctcctcgtcagcttagtcgctgbcctccctacc 776
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db 1351 yyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyy 1410

QY 777 ggccttagccttaattacctgaagaattccaggaacagctagctctcctagctagtc 836
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db 1411 yyyyyyyyyyyyyyyyyyyyyyygtagcCAAAATCTCTACTCTTTACACTACTTGCA 1470

QY 837 taaccttaatgcaatcagc 856
      | ||| | | | | | |
Db 1471 GATAGGTAAATTCAGCTGATG 1490

```

RESULT 2  
 US-08-232-463-14/c  
 Sequence 14, Application US/08232463  
 Patent No. 5670367  
 GENERAL INFORMATION:  
 APPLICANT: DORNER, F.  
 APPLICANT: SCHEIFLINGER, F.  
 APPLICANT: FALKNER, F. G.  
 TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
 NUMBER OF SEQUENCES: 52  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Foley & Lardner  
 STREET: 1800 Diagonal Road, Suite 500  
 CITY: Alexandria  
 STATE: VA  
 COUNTRY: USA  
 ZIP: 22313-0299  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/232,463  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/07/935,313  
 FILING DATE:  
 APPLICATION NUMBER: EP 91 114 300.6  
 FILING DATE: 26-AUG-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: BENT, Stephen A.  
 REGISTRATION NUMBER: 29,768  
 REFERENCE/DOCKET NUMBER: 30472/114 IMMCO  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703)836-9300  
 TELEFAX: (703)683-4109  
 TELEX: 899149  
 INFORMATION FOR SEQ. ID NO: 14:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 7218 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single

```

; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PTZgpt-F1s
US-08-232-463-14

```

|    |             |  |              |       |            |        |        |        |       |    |
|----|-------------|--|--------------|-------|------------|--------|--------|--------|-------|----|
|    | Query Match | Similarity   | 8.3%         | Score | 38.6       | DB     | 1,     | Length | 7218: |    |
|    | Best Local  | Similarity   | 8.0%         | Pred. | No. 0      | 0.034: |        |        |       |    |
|    | Matches     | 20;  | Conservative | 131;  | Mismatches | 100;   | Indels | 0;     | Gaps  | 0; |
| QY | 273         | aggggctgcacagaagccacaggaagcagcgccaggtcttgagagatggtcccaagaag  | 332          |       |            |        |        |        |       |    |
| Db | 1276        | RRR    | 1217         |       |            |        |        |        |       |    |
| QY | 333         | cggccaagaatgtagtgcgaagattggtcttcctgagagcccgcgaagaataatcatgac | 392          |       |            |        |        |        |       |    |
| Db | 1216        | RRR    | 1157         |       |            |        |        |        |       |    |
| QY | 393         | agtgtctggtgcgaagaagcagtcgccctgtaattcaaggacaatgtgaagaa        | 452          |       |            |        |        |        |       |    |
| Db | 1156        | RRR    | 1097         |       |            |        |        |        |       |    |
| QY | 453         | aacaagacaccaagaccacacagaagaacgaacagcattccagagatgccaggaat     | 512          |       |            |        |        |        |       |    |
| Db | 1096        | RRRRRRRRRRRRRRRRRRRRRRRRRRRRATCGAAGCTCCCTGACCTGCAGCAAGC      | 1037         |       |            |        |        |        |       |    |
| QY | 513         | tctcaacaat   | 523          |       |            |        |        |        |       |    |
| Db | 1036        | TTCGGATTAAAT   | 1026         |       |            |        |        |        |       |    |

```

RESULT      3
US-09-007-005-17/c
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Robert, Richard W.
; APPLICANT: Liu, Rihue
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; TITLE OF INVENTION: FUSIONS
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: 05/09/007, 005B
; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035, 963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064, 491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ. ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
; US-09-007-005-17

```

Query Match 3.8%, Score 33.8; DB 4; Length 289;  
Best Local Similarity 5.3%; Pred.No. 0.2;  
Matches 11; Conservative 91; Mismatches 107; Indels 0; Gaps 0;

Qy 624 cactctctctccaccactcactccacatgataccaccctaattcagtcgc 683  
:: : : : : : : : : : : : : : : :  
Db 229 YALGCTTGGYGVNNYNINYSNINYSNINYSNINYSNINYSNINYS 170

Qy 684 tcaaaagcatgttttcaagaatatttgttgctctctclagtctctctct 743  
: : : : : : : : : : : : : : : :

```

Db      169 YNUNSYNNYNNYSYNNSYNNSYNNSYNNSYNNSYNNSYNNSYNNSYNNSYNNS 110
Oy      744 cgtcagcttcaggccctgccccctaccgaagcttaggcctaattacccgaagaat 803
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      109 YNNNSYNNSYNNSYNNSYNNSYNNSYNNSYNNSYNNSYNNSYNNSYNNSYNNS 50
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      804 ccaggaactgtagcttacctagctagt 832
          : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      49 YGYTAAVAYTYTGTTAATAVATTAAGY 21

RESULT      4
US-09-244-796-17/C
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihue
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; EARLIER FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Translation template
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

Query Match      3.8%, Score 33.8; DB 4; Length 289;
Best Local Similarity 5.3%; Pred. No. 0.2;
Matches 11; Conservative 91; Mismatches 107; Indels 0; Gaps 0;

Oy      624 cactctttctcccacccaactctcccaactgtaaccacccctaataatccagtgctc 683
          :|: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      229 YAYGCVNYGVSYNNNSYNNSYNNSYNNSYNNSYNNSYNNSYNNSYNNSYNNS 170
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      684 tcaaaagcatgtttttaagaatcatttgtgtgtgtgtctctctctagtgcttctctct 743
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      169 YNNNSYNNSYNNSYNNSYNNSYNNSYNNSYNNSYNNSYNNSYNNSYNNSYNNS 110
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      744 cgtcagcttcaggccctgccccctaccgaagcttaggcctaattacccgaagaat 803
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      109 YNNNSYNNSYNNSYNNSYNNSYNNSYNNSYNNSYNNSYNNSYNNSYNNSYNNS 50
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      804 ccaggaactgtagcttacctagctagt 832
          : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      49 YGYTAAVAYTYTGTTAATAVATTAAGY 21

RESULT      5
US-09-247-155-53/C
; Sequence 53, Application US/09247155A
; Patent No. 6312922
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclet, Aymeric

```

```

; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: Complementary DNAs
; FILE REFERENCE: GENSET.021A
; CURRENT APPLICATION NUMBER: US/09/247,155A
; CURRENT FILING DATE: 1999-02-09
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-02-09
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-08-10
; EARLIER APPLICATION NUMBER: 60/099,273
; EARLIER FILING DATE: 1998-10-04
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: Patent.pm
; SEQ ID NO 53
; LENGTH: 445
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 81..356
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 81..152
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 6.2
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 406..411
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 429..445
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1
; OTHER INFORMATION: n=a, g, c or t
; US-09-247-155-53

Query Match          3.7%; Score 32.8; DB 4; Length 445;
Best Local Similarity 53.0%; Pred.No.0.54;
Matches 70; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 207 aatctctccctccctcgtgtgctgcacataatgctgcatgctcctagagcct 266
      ||| || ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 333 aatcggctggggcctctcttatcttgcttgcagctgcgctggcgctcctccagctat 274

QY 267 gatccaggggtcgcacagagccacagggagccgagggcctcttagagagctcca 326
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 273 GAAAGGTGGCCACACAGATGAGCCACAGTGGCCAGGGGATGAGCGCGGCGACACGAGA 214

QY 327 ggaagggcgcca 338
      ||||| ||
DB 213 ACTAGGCGGCGCA 202

RESULT 6
US-09-228-986-3
; Sequence 3, Application US/09228986
; Patent No. 6359198
; GENERAL INFORMATION:
; APPLICANT: Strubala, Timothy
; APPLICANT: Neuenhuizen, Niels
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signalling
; FILE REFERENCE: 11000/1020
; CURRENT APPLICATION NUMBER: US/09/228, 986
; CURRENT FILING DATE: 1999-01-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3

```

```

RESULT 7
US-08-458-568A-11
: Sequence 11, Application US/08458568A
: Patent No. 5821339
: GENERAL INFORMATION:
: APPLICANT: Schaffer, Priscilla A.
: APPLICANT: Yeh, Lily
: TITLE OF INVENTION: Compositions and Methods for Treatment of Herpesvirus
: TITLE OF INVENTION: Infections
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339r1s
: STREET: One Liberty Place, 46th floor
: City: Philadelphia
: STATE: PA
: COUNTRY: USA
: ZIP: 19103
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Wordperfect 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/458,568A
: FILING DATE: 02-JUNE-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/065,146
: FILING DATE: 05-MAY-1993
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Leary Ph.D., Kathryn R.
: REGISTRATION NUMBER: 36,317
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 568-3100
: TELEFAX: (215) 568-3439
: INFORMATION FOR SEQ ID NO: 11:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 12001 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double

```

RESULT 8  
 US-08-931-999-4  
 : Sequence 4: Application US/08931999  
 : Patent No. 6043219  
 :  
 : GENERAL INFORMATION:  
 : APPLICANT: Iandolo, John J.  
 : APPLICANT: Crupper, Scott S.  
 : TITLE OF INVENTION: Broad Spectrum Chemotherapeutic Peptides  
 : NUMBER OF SEQUENCES: 4  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Hovey, Williams, Timmons & Collins  
 : STREET: 2405 Grand Boulevard, Suite 400  
 : CITY: Kansas City  
 : STATE: Missouri  
 : COUNTRY: U.S.A.  
 : ZIP: 64108  
 :  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: Patent Release #1.0, Version #1.30  
 :  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/931,999  
 : FILING DATE:  
 : CLASSIFICATION: 514  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: US 08/710,561  
 : FILING DATE: 19-SEP-1996  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Collins, John M.  
 : REGISTRATION NUMBER: 26,262  
 : REFERENCE/DOCKET NUMBER: 25043-A  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: 816/474-9050  
 : TELEFAX: 816/474-9057  
 : INFORMATION FOR SEQ. ID NO.: 4:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 6755 base pairs  
 : TYPE: nucleic acid  
 : STRANDEDNESS: double  
 : TOPOLOGY: unknown  
 : MOLECULE TYPE: DNA (genomic)  
 : HYPOTHETICAL: NO  
 : ANTI-SENSE: NO  
 : ORIGINAL SOURCE:  
 : ORGANISM: Staphylococcus aureus



QY 439 ggcgaatgttgaaagaaaacagaacaccaaggca 470  
- - - - -  
Db 330 TCACACATGACATTACCAATACTCCAAAGCA 361

RESULT 11

US-08-592-126-62/c/  
Sequence 62, Application us/08592126  
Patent No. 5821091  
GENERAL INFORMATION:  
APPLICANT: Gregory Dolganov  
TITLE OF INVENTION: Transcripts Encoding Immunomodulatory  
TITLE OF INVENTION: Polypeptides  
NUMBER OF SEQUENCES: 151  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: 350 Cambridge Avenue, Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306

|                       |                 |                |           |              |
|-----------------------|-----------------|----------------|-----------|--------------|
| Query Match           | 3.48;           | Score 30.8;    | DB 1;     | Length 1698; |
| Best Local Similarity | 54.48;          | Pred. No. 5.2; |           |              |
| Matches 62;           | Conservative 0; | Mismatches 52; | Indels 0; | Gaps 0;      |

|    |      |  |      |
|----|------|--|------|
| OY | 215  | cctcctccctgttgcgccaactaatgctgatgtcatcagtccttagcacgctga   | 268  |
|    |      |  |      |
|    |      |  |      |
| Db | 1318 | ccatctttccccttggtgccacgacctgctgaaccggatgtacgacaggattctca | 1265 |

## RESULT 12

```

US-08-073-383-5/c
; Sequence 5, Application US/08073383
; Patent No. 5443962
;
; GENERAL INFORMATION:
;
; APPLICANT: Diaretta, Giulio
; APPLICANT: Collarel, Guillaume
; APPLICANT: Damaquez, Veronique
; TITLE OF INVENTION: (fill in "insert title of Application" class and reagents for

```

```

; TITLE OF INVENTION: ANTI-PROLIFERATIVE AGENTS
;
; NUMBER OF SEQUENCES: 6
;
; CORRESPONDENCE ADDRESS:
;
; ADDRESS:
;

```

|                       |                 |                |           |              |
|-----------------------|-----------------|----------------|-----------|--------------|
| Query Match           | 3.48;           | Score 30.8;    | DB 1;     | Length 2062; |
| Best Local Similarity | 58.9%;          | Pred. No. 5.8; |           |              |
| Matches 53;           | Conservative 0; | Mismatches 37; | Indels 0; | Gaps 0;      |

|    |     |                               |     |
|----|-----|-------------------------------|-----|
| QY | 733 | tctctctctcgtcagctcttagcctgtgc | 762 |
|    |     |                               |     |
| Db | 902 | TTTTCCACCTGCTTCAGTCTTGCCCTGTC | 873 |

### RESULT 13

US-08-328-239A-3/C  
Sequence 3, Application US/08328239A  
Patent No. 6037136  
GENERAL INFORMATION:  
APPLICANT: Beach, David H.  
APPLICANT: Galatimov, Konstantin  
APPLICANT: Jessus, Catherine  
TITLE OF INVENTION: Interactions between Raf Proto-Oncogenes  
TITLE OF INVENTION: and CDC25 Phosphatases, and Uses Related Thereto  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY, HOAG & ELIOT  
STREET: One Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII(text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/328,239A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: CSV002.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 832-1000
TELEFAX: (617) 832-7000
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2062 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 211..1631
US-08-328-239A-3

```

Search completed: September 16, 2002, 03:27:01  
Job time: 9344 sec

---